UNIVERSITY OF CANTERBURY

Hand gesture-based interaction in an immersive cinematic environment

Author:
Humayun Khan

Senior Supervisor:
Prof. Rob Lindeman

Co-Supervisor:
Dr. Gun Lee

Associate Supervisor:
Prof. Mark Billinghurst

A thesis submitted in fulfillment of the requirements for the degree of Master of Human Interface Technology

in the

HIT Lab NZ
College of Engineering

March 2017

Abstract

With cheaper head-mounted displays (HMDs) available in the consumer market, new virtual reality (VR) applications are being explored, one such being interactive VR movies. In conventional movies, the user passively watches a movie on a 2D screen and the level of immersion depends on the narrative of the story. In VR, the user's field of view is covered by the digital content, thus interaction with the digital content can affect the user's immersion in the virtual environment. This thesis explores how hand gesture-based interaction and hand appearance affect the user's immersion and embodiment in a 360 VR movie.

A prototype was developed by combining a Leap Motion controller and a SoftKinetic RGB-D camera. It captures the user's real hands and blends them into the virtual environment. The prototype also supports natural interaction with the virtual environment. A user experiment was conducted using the prototype to investigate the effect of hand-based interaction and hand appearance on the user's immersion and embodiment. There are two conditions for the hand appearance: the real hand, and the virtual hand. There are also two conditions for the hand-based interaction: with interaction, and without interaction.

Results showed that the real hand increased the user's embodiment in a 360 VR movie. However, hand appearance did not have any effect on the user's presence. It was found that having hand-based interaction increases the user's embodiment in the 360 VR movie. Similarly, it also displayed no effect on the user's presence. User feedback collected from the study identifies the limitations of the study and the system.

Acknowledgements

This thesis would not have been possible without the support and guidance of my senior supervisor Prof. Rob Lindeman and co-supervisor Dr. Gun Lee. Their advice was instrumental in conducting this research. I am also thankful to Ken Beckman for his help and encouragement through out the study. I would also like to thank HITLabNZ students and staff specially Hyungon Kim, Ser Qi Min Joshua Chen and Jackie Soady for their support and encouragement through the difficult times. I am really grateful to MBIE for funding my research as a part of HDI^{2}4D project. I would also like to thank my family for their constant love and support.

Contents

List of Figures

List of Tables

Abbreviations

VR	**V**irtual **R**eality
MR	**M**ixed **R**eality
AV	**A**ugmented **V**irtuality
AR	**A**ugmented **R**eality
HMD	**H**ead-**M**ounted **D**isplay
CG	**C**omputer**G**raphics
RGB	**R**ed **G**reen **B**lue
GPU	**G**raphic **P**rocessing **U**nit
API	**A**pplication **P**rogramming **I**nterface
SDK	**S**oftware **D**evelopment **K**it
SVD	**S**ingular **V**alue **D**ecomposition

Chapter 1

Introduction

1.1 Background and Motivation

The movie industry has been developing together with advance in technology, such as computer graphics, digital image and video processing, and spatial audio. Recently with advance in virtual reality technology, the movie industry is facing another tipping point where the conventional ways of recording and presenting need to be re-investigated. The assumption that the cinematographer controls the viewer's perspective does not hold for immersive virtual reality (VR). In VR, the viewer is the master of his or her own perspective; s/he can look in any direction, and perhaps even interact with the scene. With this freedom introduced by the technology, there are many questions arising on how the movie content has to be produced. What should the viewers be allowed to do to prevent them from missing any part of the story? Should the story be linearly structured, or should it change with the viewer's actions? If a group of friends wants to watch it together, how would the experience be? When do animated movies offer better experiences than live-action movies, and vice versa? Some of these questions will take years to solve, and more questions will arise along the way. The arrival of affordable high quality head-mounted displays (HMDs) has accelerated the process of answering these research questions.

Among various different VR techniques applicable, 360 degree panorama video is most actively adopted to movie production. In 360 degree movies, the viewer is completely surrounded by a spherical video background, and s/he has control over the perspective in the spherical space. There could be various ways to improve movie watching experience in 360 degree movies. For example, prior work [9] has investigated how does showing the viewer's body inside the 360 movie affect on movie watching experience. Taking this to the next step, this thesis explores how adding interactivity to a 360 VR cinematic environment improves the user's sense of presence and embodiment. It also examines whether interacting with the cinematic environment using realistic computer graphics (CG) hands differ from viewing your own real hands infused into the virtual environment.

Interactivity in a movie narrative can be added with input controllers such as the HTC Vive controller[1] and Oculus Touch controller[2] or with a non-invasive technique which takes the user's hands as an input such as the Leap Motion controller[3]. Each input method has its own pros and cons and is applicable for different scenarios. In our case, we are designing the application for a home entertainment system. Our potential user, as shown in Fig. 1.1, is sitting or in a comfortable state. Therefore, we believe that a non-invasive technique is more appropriate.

FIGURE 1.1: User interacting in a movie inside the VR

To use the non-invasive technique, user's hands have to be taken as a input. A human hand is an incredibly dexterous organ, which interact and communicate with surrounding environment using a multitude of ways. To capture this dexterity, we need a recognition process that can detect, track and recognise the hands [16].

Detection is the first step in the recognition process. During detection, the hands are captured and separated from the background. They are then further segmented into

[1]HTC Controller: `https://www.vive.com/us/accessory/controller/`
[2]Oculus Touch Controller: `https://www.oculus.com/rift/`
[3]Leap Motion Controller: `https://www.leapmotion.com/`

different regions and areas for tracking. Tracking is challenging as the hand can move at varying speeds with varying speeds and its appearance can change from one frame to another. Through keeping track of the detected and segmented hand regions from one frame to another, we are comprehending the overall observed movement. The final step in the recognition process involves interpreting the hand/hands posture, location or series of movement into interaction in the virtual world. Using these three steps, hand gestures can be recognised and translated into a meaningful interaction in the virtual world.

Once a hand gesture is translated into an interaction in virtual world, it can still be interpreted in many different ways. It can have the same interpretation as the real world, for example grasping virtual object by holding, or it could have completely different meaning, for example an open hand could mean to fly the virtual object. It depends on the designed system. If this interaction is interpreted similarly as in the real world, it could reduce the cognitive load using the system. However, including the interaction in the virtual world which mimics the interaction in the real world is a complex problem and is extensively researched.

To define the problem, researchers have classified interactions in different modes. One such definition is by Mine who classified interactions into five modes [17]. They are manipulation, selection, movement, and scaling, and the last one (which is derived from first four) is virtual menu-based interaction. Manipulation is analogous to real world interaction, and it involves grabbing the virtual object using hands and physically moving it to the desired location. It can also be used to modify size, scale and other characteristics of the object [18]. The second mode is selection which involves assigning an ID to each selectable object then choosing it by using a mechanism such as a gesture or device. The third mode is movement or navigation in the virtual environment which could be flying [19] or walking in virtual space [20]. The fourth mode is scaling which allows user to increase or decrease the size of a virtual object. The fifth mode is menu-based interaction, where the user interacts with the virtual world using menus available in 3D virtual space.

From the above mentioned interaction modes, only manipulation was used for our system. Because including more than required will further complicate the interaction.

However, for creating rich cinematic experiences, different modes or a combination of modes can be used depending on the requirements of the situation.

Creating interaction with film content is a challenging task. It involves addressing complex problems in computer graphics such as estimating 3D geometry, lighting and other scene information. Supporting interaction with user's actual rendered hands further complicates the problem.

In our work, we created a mixed reality (MR) system which allows the user to transition between reality and virtuality. Similar work in the past had created MR systems that balanced the immersion in a virtual world while looking into the physical world [5]. Another similar system developed by Carrozzino et al. created an immersive virtual environment to educate users on the craft of printmaking [1]. Their system design can be seen in Fig. 1.2.

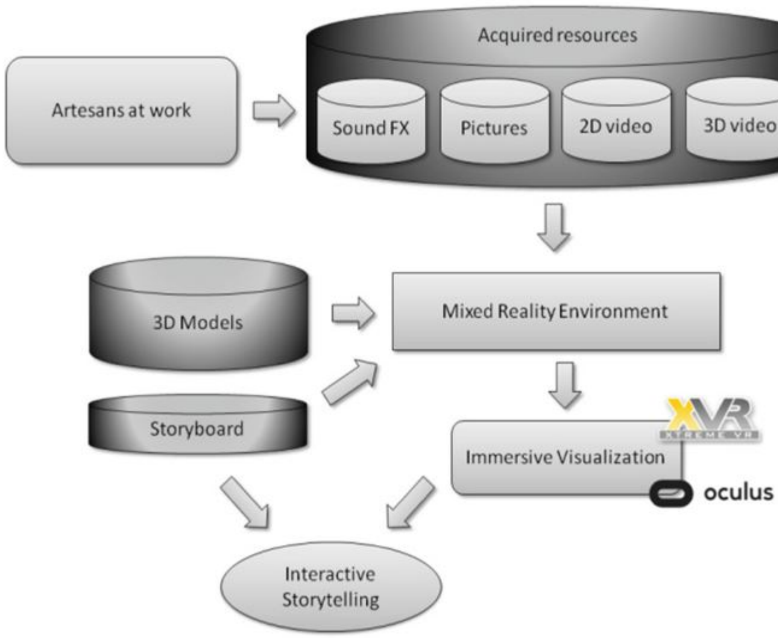


FIGURE 1.2: AMICA system design [1]

In our system, the user experiences the cinematic virtual environment on a consumer level HMD (Oculus DK2). With tracked hands, the user will also be able to interact with objects in the virtual movie scene.

1.2 Research Questions

Using the designed system, we aim to answer these research questions:

- Does hand gesture-based interaction in a cinematic virtual environment improves a users sense of presence and embodiment?
- Does the type of hand appearance (actual hands or CG hands) in a cinematic virtual environment affect the sense of presence and embodiment?

1.3 Contribution

The main contributions of this thesis are:

- A novel method to provide hand interaction in 360 degree movies by combining the Leap Motion controller and the SoftKinetic, DS325 camera. The combination used Leap Motion's hand tracking and SoftKinetic's hand visualisation technologies.
- A user study researching the effect of interaction and representation of user's body on the sense of presence and embodiment.

1.4 Thesis structure

The structure of the thesis is as follows.

Chapter 2: Discusses the related work done on mixed reality-based cinematic environment, hand-based augmented virtuality and presence and embodiment in VR.

Chapter 3: Presents the design process of the prototype.

Chapter 4: Explains the implementation of the prototype.

Chapter 5: Describes the user experiment in detail and discusses each step of the evaluation process.

Chapter 6: Presents the results obtained from the user study.

Chapter 7: Discusses the results found and describes the limitations of the study.

Chapter 8: Concludes the thesis and identifies possible future areas of research.

Chapter 2

Related Work

As mentioned in Chapter 1, this thesis aims to investigate how adding a hand gesture-based interaction to 360 degree movie watching affects the user's sense of presence and embodiment. This chapter reviews the related work done on the topic. The first section discusses the mixed reality-based cinematic environment, the second section outlines hand-based augmented virtuality, the third section is on presence and embodiment in VR and lastly, the chapter identifies research opportunities.

2.1 Mixed Reality in a Cinematic Environment

There are few research works which explore mixed reality for cinematic experiences. Cheok et al. [2] created an interactive theatre system using an external camera system consisting of 15 cameras and a HMD. The system captured the user in real-time and rendered him/her in a mixed reality environment where s/he can interact with a virtual environment. The system supported three modes of interactions. The first mode was an outdoor theatre land exploration mode which allowed users to walk around in an outdoor environment wearing a HMD (Fig. 2.1a). The virtual world was overlaid on the physical world and was seen through the first person perspective in a HMD (Fig. 2.1b). The second mode was an AR theatre land exploration mode which embedded a virtual theatre to the physical environment by merging the AR and VR world (Fig. 2.2). It also supported the interaction with the virtual objects. The third mode was a virtual interactive theatre mode which featured a fully immersive cinematic environment. A live actor in the physical world (Fig. 2.3a) was captured by the external camera system and embedded in the virtual world in real-time (Fig. 2.3b).

(a) HMD's outdoor interface

(b) Virtual actors overlaid on the real world

FIGURE 2.1: Outdoor theatre land exploration mode [2]

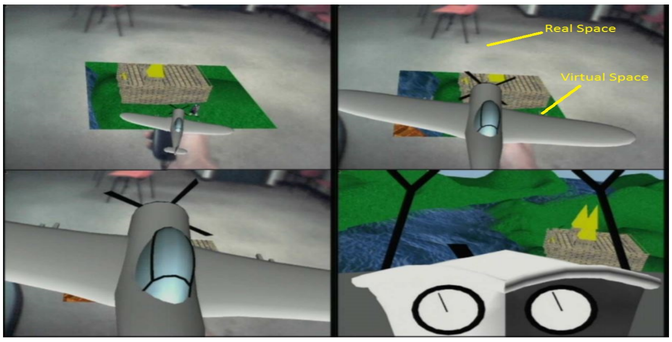


FIGURE 2.2: AR theatre land exploration mode [2]

(a) Real-time capture of the live actor

(b) HMD view of the live actor

FIGURE 2.3: Virtual interactive theatre mode [2]

The mixed reality system [2] is a conceptually unique idea, but requires an external camera system (15 cameras) which is not viable for a home entertainment system.

In other work, Tenmoku et al. [21] and Ichikari et al. [22] proposed a work flow to insert the computer graphics animation data, the motion capture data and the 3D video data into a special movie-making software. The position and movement of a camera in the real world can be incorporated into a 3D model representing the physical location of the filming. Using the 3D model of the scenery and the animation data, the movie director is able to plan ahead the movements and positions of the camera for the shot to be filmed. These systems can be used to create an offline mixed reality cinematic environment but would not be useful for a cinematic environment which is updated in real-time.

With a wider availability of cheap HMD devices, there are some VR studios which have developed immersive cinematic experiences. One example is the Frame Store VR studio [3], which developed an immersive cinematic experience based on the movie “Interstellar” [23]. In the experience, the user is sitting in a space shuttle. There are some scenes where the user experiences weightlessness which is simulated by a custom rigged chair, as shown in Fig. 2.4. The VR system offers an immersive experience but it cannot be implemented as a home entertainment system as it require special hardware (the custom rigged chair). The cost and process of installing these special hardware is prohibitive for an average home entertainment user. Therefore, creating an immersive system which offers an interactive cinematic experience could potentially be competitive. It surely would not offer a similar level of immersion, but it can go closer depending on the plausibility of the designed immersive experience.

FIGURE 2.4: Custom rigged chair to simulate weightlessness [3]

2.2 Hand-based Augmented Virtuality

There is a continuum of possible combination between the real and virtual world, ranging from superimposing virtual objects into the real world, Augmented Reality (AR), to overlaying real world objects on the virtual world, Augmented Virtuality (AV) [24]. Mixed reality encompasses both AR and AV. The continuum can be seen in Fig. 2.5. AV has been mostly used in the teleconferencing systems which integrated real world 2D video streams to a shared virtual environment [25, 26].

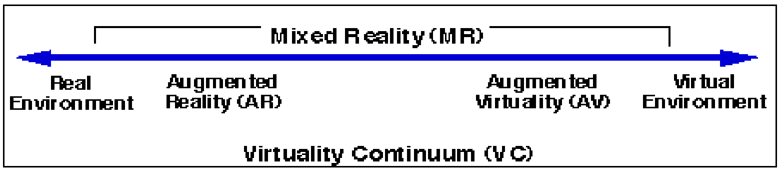


FIGURE 2.5: Real to virtual environment, virtuality continuum [4]

In one of the early works on AV, Metzger [27] used an egocentric camera as a portal to the real world by streaming video from camera to a HMD . In later works, AV has been used to blend part of the body (mostly hands) or external objects in the virtual environment [5, 6, 28, 29]. In more recent works, AV has been used to blend hands in the virtual reality as well as perform interactions with virtual environments [7, 30]. There have also been prior work which track hands using an egocentric camera and drive a CG hand model to interact with the virtual environment [14, 15]. Our research is centred on hand incorporation in VR, so we will review past research on this topic.

2.2.1 Blending hands in VR for visualisation

Bruder et. al [29] blended the user's actual hands into VR and studied its effects on sense of presence. User's hands were segmented using chroma keying with an egocentric camera and displayed in a HMD in real-time. Their experiment compared two conditions: with or without the hand visualisation. The study used the Slater-Usoh-Steed (SUS) questionnaire to measure presence [31]. Results indicated that users felt a higher sense of presence on seeing their hands in the virtual environment.

Another work by McGill et al. [5] studied the effects of interaction with real objects and the peripherals on user's sense of presence while using a HMD. Chroma keying method was used to segment and blend the hands and periphery into the virtual reality (Fig. 2.6).

This study compared interaction in the selective reality (partial or minimal blending) with the interaction in full view of the reality. The Igroup Presence Questionnaire (IPQ) was used for measuring user's sense of presence [32]. The results indicated a higher sense of presence for partial reality than the full view of reality.

Figure 2.6: left: Minimal blending (reality around users hands), middle: Partial blending (all interactive objects). right: Full view of reality [5]

Similar work by Budhiraja et al. [6] studied user's preference on showing hands and a peripheral real world in a virtual world. The study used a stereo camera setup (two Logitech C310 webcams) to segment the user's hands and the periphery. Comparisons between four conditions as shown in Fig. 2.7 found that the users prefer to see their hands, peripheral object and context of the real world in a virtual environment. Chroma keying is a common technique for the inclusion of user's hands in VR, as what these researchers have used in their studies on hand segmentation from an egocentric camera [28, 33, 34].

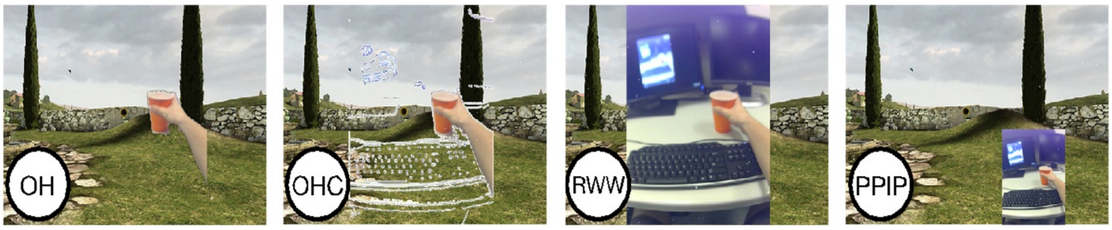


Figure 2.7: left: Object and Hand, left-middle: Object, Hand and Edges, right-middle: Real World Windowed, right: Physical Picture In Picture [6]

2.2.2 Blending hands in VR with Visualisation and Interaction

Tecchia et al. [7] explored a mixed reality rendering technique based on an egocentric camera that incorporates hands into VR and allows interaction with virtual objects. The study hypothesised that the introduction of a photo-realistic user's hand into a virtual environment induces a strong sense of embodiment without requiring a virtual avatar. A head-mounted RGBD camera was used to segment and reconstruct a 3D mesh of user's hands in the virtual environment. Their setup is shown in Fig. 2.8. Finger tracking was done by using two colour rings (blue and green) which were placed on the index finger and the thumb. The interaction with virtual objects was based on these tracked fingers. Since only two fingers were tracked, interaction was limited to only a few gestures with the virtual environment.

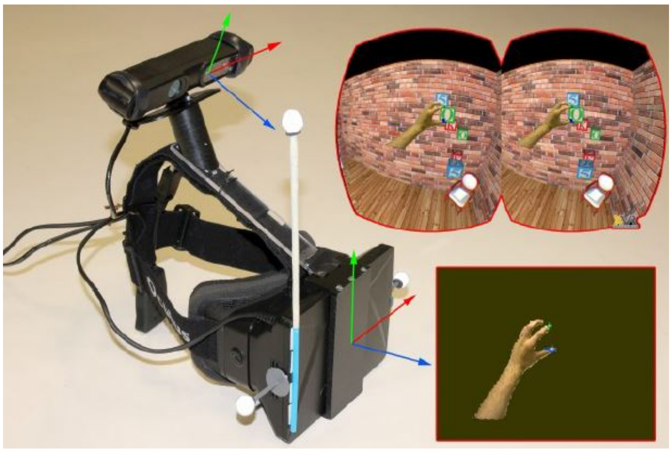


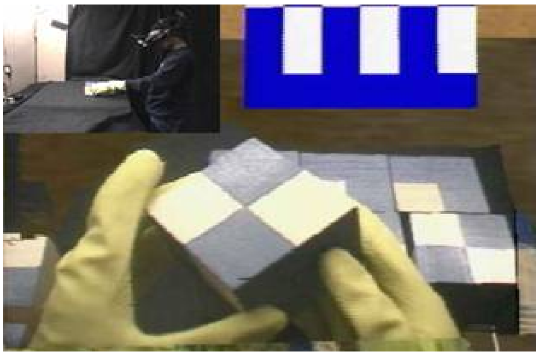
FIGURE 2.8: HMD with an egocentric RGBD camera, the coordinate reference frames of Oculus Rift and the RGBD camera are also illustrated [7]

The WeARHand method by Ha et al. [35] combined two depth cameras (near and far range) as an egocentric input for visual feedback and hand interaction. By combining two cameras the range of interaction was extended; however, it was limited to three gestures based on the finger-count method. The system also did not have physics-based interactions as there was no physics-based hand collider. Similarly, this method by Tecchia et al. [7] also had a limited amount of interactivity. Using one of these methods would result in a limited amount of interaction with the virtual environment. As our application is meant for a cinematic environment, more interactivity is required. Therefore, we looked into a commercially available solution for hand tracking which is discussed in the next chapter.

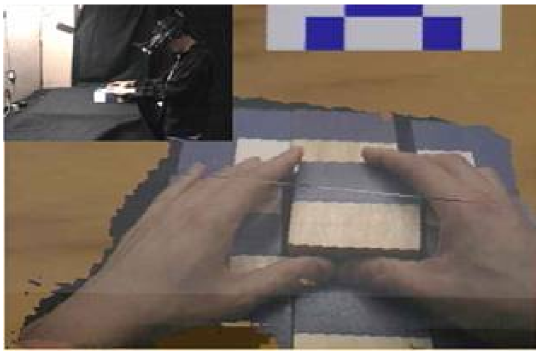
2.3 Virtual Avatar, Sense of Presence and Embodiment

Presence in the VR is commonly confused with immersion. Slater and Wilbur [36] made a clear distinction by defining presence as “a state of consciousness, the psychological sense of being in the virtual environment”, and immersion as “a description of a technology, and describes the extent to which the computer displays are capable of delivering an inclusive, extensive, surrounding, and vivid illusion of reality to the senses of a human participant". A research conducted by Lok et al. [8] studied the effects of self-avatar fidelity on the task performance and user's sense of presence. The study had three different self-avatar representations which are shown in Fig. 2.9.

(a) Virtual blocks and generic avatar



(b) Real blocks and generic avatar



(c) Real blocks and visually faithful avatar

FIGURE 2.9: Three different self-avatar representations [8]

The Slater-Usoh-Steed (SUS) questionnaire was used to measure presence [31]. The quantitative data showed no significant difference for self-avatar fidelity. However, the qualitative feedback indicated that the users felt higher sense of presence with higher fidelity of the self-avatar. In their discussion, Lok et al. also indicated that there might be a difference on sense of presence with the higher fidelity of self-avatar. In another work, Usoh et al. [37] hypothesised on the visual fidelity of avatar, "substantial potential presence gains can be had from tracking all limbs and customising avatar appearance". These works shows that there can be a potential increase in presence by including interaction and higher fidelity (more realistic) self-avatars.

The sense of embodiment is defined by Kilteni et al. [38] as "SoE toward a body B is the sense that emerges when Bs properties are processed as if they were the properties of ones own biological body". They further divided embodiment into three parts: sense of self-location, sense of agency, and sense of body ownership. We will briefly define these three parts. The self-location can be defined as the space in which a person perceives themselves to be located. Agency is the feeling that a person is the agent of his/her own actions. Agency is closely related to the awareness of one's own action [39]. Ownership is the feeling that ones own body is the source of sensations [40]. Our project is focused on only two dimensions of embodiment: agency, and ownership. We will briefly discuss some of the related work on agency and ownership.

Lin et al. studied the effects of hand appearance on the sense of agency and ownership. In the study, users were shown six hand models: realistic hand, toony hand, very toony hand, zombie hand, robot hands and wooden block. The study found no effect of hand appearance on the sense of agency. The sense of agency remained intact even for non-realistic avatars [14]. More realistic hands showed stronger sense of ownership. Another work by Argelaguet et al. [15] studied the sense of agency and ownership by having interaction in a virtual environment using different hand model representations. Three hand models were used: realistic, iconic and abstract hand model. They found that the sense of agency is related to the virtual hand control and efficiency of performing a task in a virtual environment, and the sense of ownership is mainly related to the visual appearance of the virtual hand. Results of the study showed that the sense of agency was stronger for abstract and iconic hand models than realistic hand models. In contrast, the sense of ownership was stronger for more realistic hands than the iconic and abstract hands. Argelaguet et al.'s results differed from Lin et al.'s results who did not find any effect of hand appearance on the sense of agency. For the sense of ownership, results from both studies were similar.

Slater et al. [41] conducted a study to understand what induces an illusion of ownership over a virtual arm. It found that the synchrony between visual and proprioceptive information along the motor activity induces the illusion. They further studied the illusion of ownership by extending the virtual arm and found that the brain still perceive it as its own even when the arm is extended three folds [38]. In another research, Slater et al. studied racial bias by inducing the illusion of ownership on a virtual body. It showed that the racial bias could be reduced by using an avatar of a black person [42].

2.4 Research Opportunities

Research has shown a strong evidence that an increase in virtual hand (or virtual avatar) fidelity improves the user's sense of ownership in the virtual environment. It has also indicated the effect of a virtual hand (or virtual avatar) appearance on user's sense of agency. In past studies, the sense of agency have been found to increase with higher interactivity. The sense of embodiment comprises of three components, and two of them are agency and ownership. Having an increase in one of the two would result in an increase in overall embodiment. Hence, there is research potential to study the embodiment with high fidelity virtual hands and increased interaction in the virtual environment.

The sense of presence has been proven to increase by showing the user's real hand or by having interactivity with the real world. In order to add interaction, the user's hand must be tracked. With currently available technology, tracked hands are used to drive a virtual hand model. Unlike real hand interactions with the virtual world, there is extensive existing research done on hand tracking, however, most of the research uses static exocentric cameras which cannot be used on a constantly moving HMD. An egocentric camera is needed to perform hand tracking. There are research which used an egocentric camera to visualise real hand and perform hand tracking, but they support limited interaction. Commercially available hand trackers offer more interactivity, but they use tracking information to drive virtual hand models and do not support real hand visualisation. Therefore, we have built a solution to combine real hand visualisation with commercially available hand tracking to create a hybrid solution which will support the user's real hand visualisation as well as better interactivity.

Cinematic storytelling in the VR is in a nascent stage. Storytellers are experimenting new ways to tell stories in an immersive manner. So far no rules have been publicly established to tell a good VR story. Therefore, this has been a good opportunity to explore this space and try new methods such as adding meaningful interaction with the virtual environment or including part of the real world. The combination of the virtual hand appearance and interaction in a cinematic environment is an under-explored research area and has the potential of creating fresh new cinematic experiences.

Chapter 3

Design Process

This chapter presents material on the design process used to create the prototype. The design process began by gathering requirements. This was followed by ideation, and finally the prototype design.

3.1 Requirements Gathering

The project goal was to develop a system which blends the user's body in a cinematic VR scene and allows user to physically interact with the objects in the scene. Blending the body involves two components: first, transporting the whole body into the cinematic environment, second, allowing hand gesture-based interaction with the cinematic environment. The developed system is meant for home users who can wear the system in a comfortable space such as sitting on the couch. The developed system would then be assessed for immersiveness through user evaluation.

Besides reviewing the project proposal, we also looked at previous work done by Chen [9]. Chen worked on blending the user's body in a 360 movie and studied the effects of the blended body on user's immersiveness. He found that blending alone does not increase immersiveness in a cinematic scene as participants felt purposeless in the cinematic VR scene without having interactions. From the qualitative data, he concluded that user immersiveness in a cinematic scene can be increased by including the meaningful interaction and by blending the user narratively. He gave a few examples of meaningful interactions, such as interaction with a virtual character as shown in Fig. 3.1. Another example he gave was using a phone as a weapon which can be seen in Fig. 3.2.

FIGURE 3.1: Interaction with virtual characters [9]

FIGURE 3.2: Using a phone as a weapon [9]

3.2 Ideation

In virtual reality, the digital environment completely surrounds the user's field of view and offers a new way to connect with the digital content. It also provides a new medium for cinematic storytelling and interactivity, which did not work for conventional cinematic mediums [43] such as TVs. There are different amounts of interactivity that can be added in a cinematic scene. For interactive VR 360 videos, Dolan and Parets [10] defined four different quadrants of interactive influence and existence in the video. The quadrants can be seen in Fig. 3.3 .

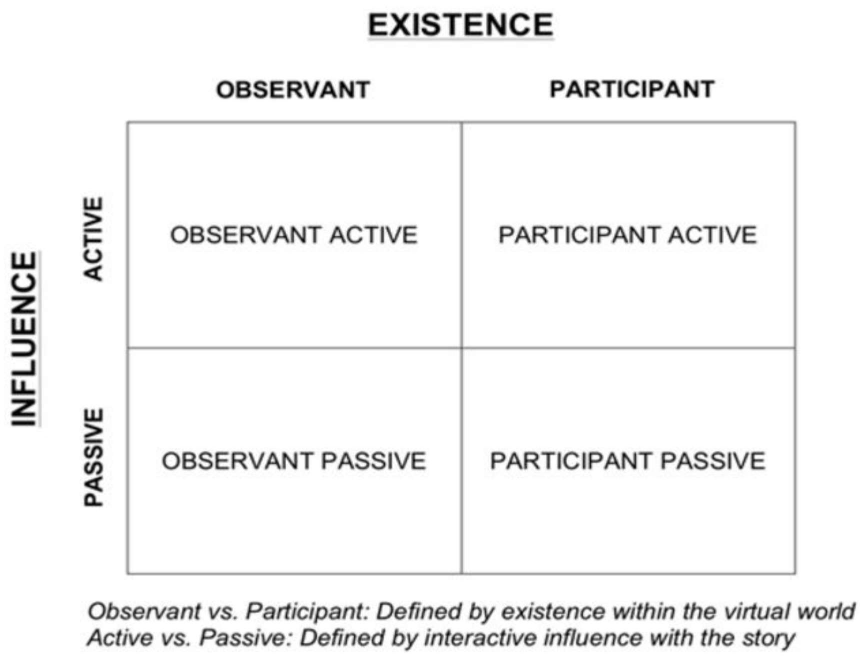


FIGURE 3.3: Four quadrants defining the extent of interaction in a VR video [10]

- Observant Passive - It is a standard model which is currently being used to watch movies.
- Observant Active - It is an interactive model which has multiple narratives for the same video, and the viewer is given a choice at a freeze frame to choose the direction of the story. An example of this model is a movie "I'm Your Man" [44].
- Participant Active - It is a model where the viewer can interact with the cinematic scene which in turn can change the narrative of the scene.
- Participant Passive - It is a model where the viewer exists as a character in the cinematic scene. They can interact with the scene but cannot influence the narrative.

For our project, we considered the participant active and participant passive roles. While exploring the participant active scenario, we came across Henry [1], a VR animation video about a hedge hog who does not have any friends and is celebrating his birthday party. The animation is created in Unreal Engine 4 by Oculus Story Studio. The Oculus Story Studio have made the Henry animation assets open source which allows us to modify the animation's narrative and study the effects of hand-based interaction on user presence.

[1]Henry, `https://www.oculus.com/story-studio/films/henry/`

We thought of different scenarios for adding interaction in the video such as lighting the candle shown in Fig. 3.4a or having another party horn on the table shown in Fig. 3.4b which user can pick-up and blow, and on blowing the horn, Henry could acknowledge the action. However, we decided not to go with the participant active model because it restricts us only to computer generated animation content, and the real life videos have to be excluded. Participant active scenarios also affect the narrative which means 360 VR video cannot be used, because in the 360 video, the narrative is fixed to the initial video. However, we can create the participant passive role in the 360 video by adding in 3D objects related to the video and allowing the user to interact with the added objects. Therefore, we decided on the participant passive model for creating interactive experiences in 360 cinematic scenes.

(a) Candle that can be lighted

(b) Henry with party horn

FIGURE 3.4: Possible interaction with Henry

3.3 Design Consideration

To include interaction in a cinematic VR scene, we need a mechanism to take user input. The input could be in the form of a controller, a wearable hand device such as glove or markers, or bare hands. As our prototype is meant for a cinematic environment, bare hand input would be more natural for users, while wearing gloves or carrying controllers could hinder natural user interaction. Including bare hands as an input requires tracking the user's hands, which is done by a hand tracker. For our prototype, we investigated three different hand trackers based on their availability and accessibility i.e. either they are open sourced or internally developed by HIT Lab NZ. The three hand trackers are:

Leap Motion controller, Robust Articulated-ICP[11], and RGBD-based hand tracker[12]. The following section will discuss them in detail.

3.3.1 Leap Motion Controller

The first hand tracker we investigated is the Leap Motion controller. The device is capable of tracking hands at up to 200 frames per second (fps) and has an effective range of approximately 2.5 cm to 80 cm from the device [45]. It has a wide field of view of about 150 degrees on the long side and 120 degrees on the short side (Fig 3.5). This allows a large interactive space for the controller of about 8 cubic feet with the shape of an inverted pyramid. The Leap Motion underlying software re-initialises the tracker as soon as it loses tracking which makes it more robust than the other trackers. It also has a smaller size which makes it easily mountable on a HMD.

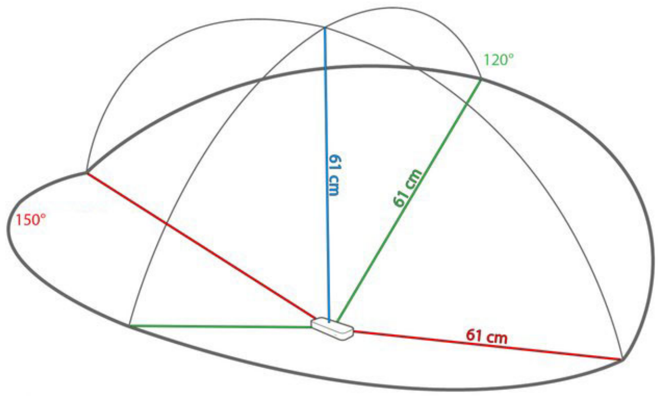


FIGURE 3.5: Leap Motion field of view

Apart from tracking capabilities, the Leap Motion also has more development resources available for the VR applications such as rigged hand models, Unity hand attachment modules, and Leap Motion SDK. With these resources, using the Leap Motion for hand tracking in VR is comparatively easier than the other hand trackers. The Leap Motion also has some limitations which is the loss of hand tracking when the two hands are too close to each other. There is also no colour (RGB) camera to show users real life hands which is required to create an augmented virtuality experience.

3.3.2 Robust Articulated-ICP for Hand Tracking

The second hand tracker that was considered for this project was an open source hand tracking algorithm which can be implemented on short range RGBD cameras. Some examples of these include Primesense Carmine and Creative Senz3D. It uses a model-based approach to infer complex hand poses by fitting 3D articulated hand model to a low-quality depth map as shown in Fig. 3.6. It can track hands in real-time for up to 120 fps. Unlike Leap Motion, it works well even when the two hands are close or clasped together. Another advantage is that it is an open source algorithm which can be modified and changed as per our requirements.

FIGURE 3.6: 3D articulated hand model from low-quality depth map [11]

We tested Robust Articulated-ICP algorithm with the Primesense Carmine 1.09 camera as shown in Fig. 3.7. It is able to track hands within the camera depth range, but if the hands go out of the range, it affects tracking and the hand starts appearing at incorrect positions as can be seen in Fig. 3.8. Another constraint with the algorithm is that for tracking to work accurately, a wristband needs to be worn.

FIGURE 3.7: Primesense Carmine 1.09 camera

FIGURE 3.8: Tracking failure cases

The main limitation of the algorithm is that it requires the RGBD camera to be fixed so it cannot be used as an egocentric camera on an HMD which keeps moving. Due to this limitation, we decided not to use the Robust Articulated-ICP algorithm for our project.

3.3.3 RGBD-based Hand Tracker

The third hand tracker we examined was developed by Bai et al. at HIT Lab NZ [12]. It is implemented using Softkinetic's DS325 RGBD camera. It segments hands from the background using skin colour and depth information. The segmentation process can be seen in Fig. 3.9. The extracted hand is further processed to retrieve the hand skeleton, as shown in Fig. 3.10a, b. The retrieved hand skeleton is then projected into a 3D space using the depth map which gives back a 3D hand skeleton (Fig. 3.10c). This 3D hand skeleton can be used to identify fingertips and palm position. With fingertips and palm position identified, we can insert a simple hand model to represent fingertips as spheres and palms as circular disks, as illustrated in Fig. 3.10d.

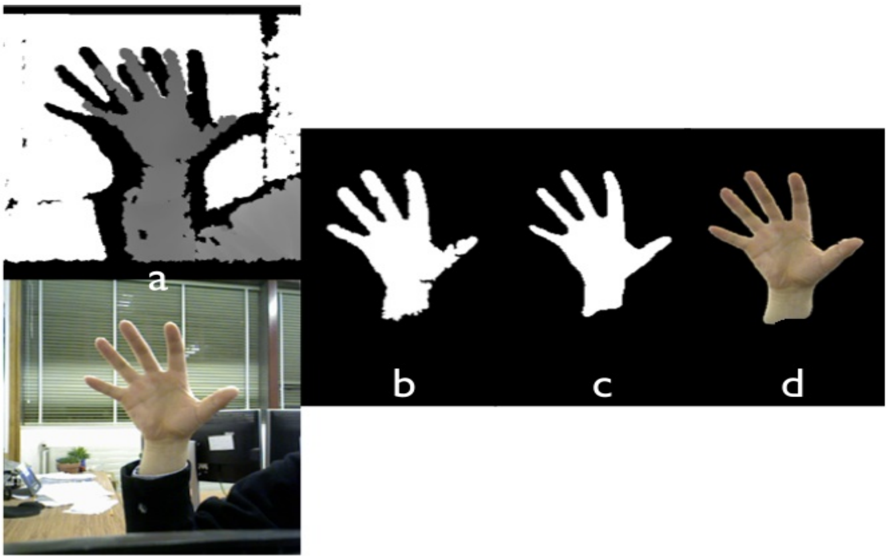


FIGURE 3.9: (a) Colour and depth frames (b) Foreground depth region (c) Smoothed hand contour (d) Final hand region.[12]

FIGURE 3.10: (a) Extracted hand contour (b) Retrieved Hand Skeleton (c) 3D Hand Skeleton (d) Simple hand model with finger tips as spheres and palm as disk [12]

The simple hand model allows limited interaction as the tracking is limited to the fingertips and palm. Limited tracking also means smaller non-continuous collider which allows restricted collision between the hands and other 3D virtual objects. The algorithm can track hands up to 30 frame per second with the Softkinetic camera. It can effectively work within the range of 6 inches to 3.3 feet from the camera. However, the tracker cannot be used in its current state for our application and has to be improved by detecting and tracking other parts of the hand, which would not be able to fit into the time frame of this thesis.

We decided to use the Leap Motion controller as the hand tracker for our project due to its better hand tracking capabilities and wider field of view. However, as the Leap Motion controller lacks a colour camera (required for real hand visualisation), we decided to combine it with the RGBD-based Hand Tracker (the third hand tracker investigated) to incorporate its tracking and colour information. The tracking information can be used to calibrate the two cameras which in turn allows colour information to be used for real hand visualisation. The calibration process is explained in the next chapter.

Chapter 4

Prototype Development

This chapter describes the implementation of a working prototype. The prototype was developed to be robust enough for experiencing an immersive cinematic environment and for carrying out a user study. It was built by combining the two hand trackers selected during the design process. The two hand trackers use Leap Motion and SoftKinetic cameras for tracking user's hand. The cameras were combined using SoftKinetic's DepthSense SDK and Leap Motion's Unity assets. Using this system real hands with textures were tracked in real-time. To play the 360 video, a 360 video player was created using the Unity sphere primitive and Easy Movie texture plugin [46]. Interactable 3D objects were also added to provide an immersive cinematic experience.

4.1 System Architecture

The system architecture of our prototype can be seen in Fig. 4.1. The diagram visually describes the whole architecture of our prototype. It describes how the image captured by the inputs (Leap Motion and SoftKinetic) goes to the output. It also describes how the 360 movie is played on the output (Oculus DK2 and headphones). It goes through all the underlying layers of processing the inputs go through before being displayed on the output.

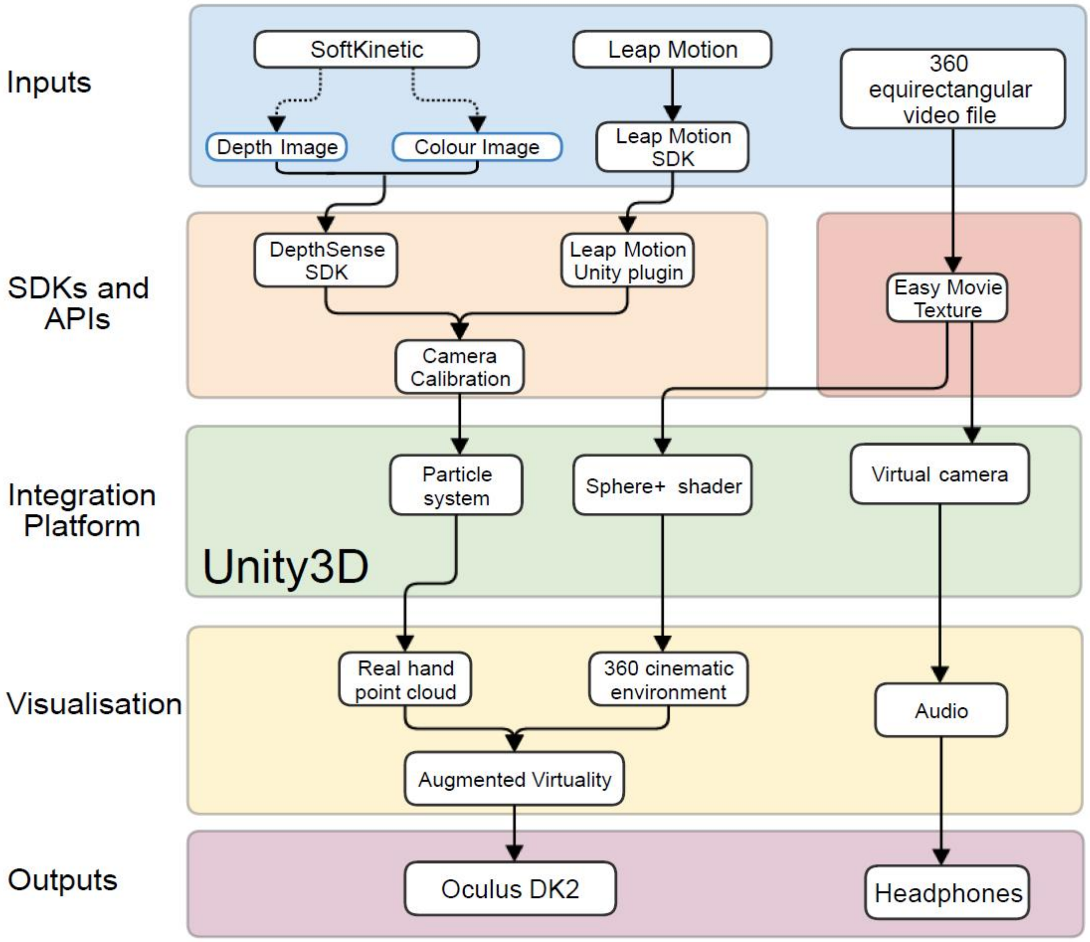


FIGURE 4.1: System Architecture

4.2 Hardware

4.2.1 Leap Motion

Leap Motion is the main hand tracker for our system. It is a motion-sensing device which uses infrared light to track hands and fingers in 3D space. The Leap Motion controller is shown in Fig. 4.2. It emits an infrared light from the three LEDs, which is reflected by the hands and is captured by the two cameras. The two streams of captured images are compared and 3D position of the hands are calculated based on that comparison.

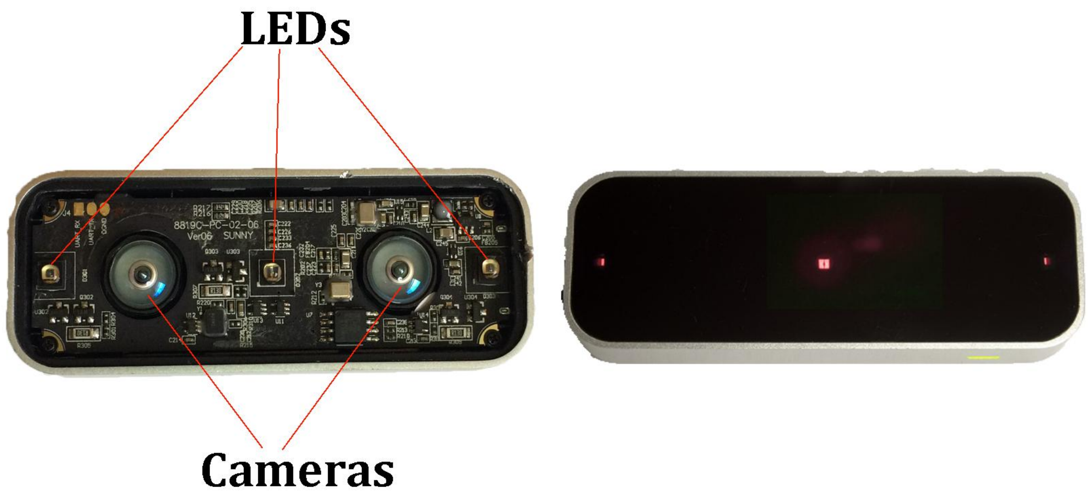


FIGURE 4.2: Leap Motion controller

4.2.2 SoftKinetic, DS325

SoftKinetic is the secondary hand tracker for our prototype. It is used mainly for adding real hand textures to the tracked hands. SoftKinetic DepthSense 325 camera is shown in Fig. 4.3. It is a time-of-flight RGB-D camera which calculates the depth information by measuring the travelling time from an infrared light emitter to the depth camera. It has a 720p RGB camera which can work upto 30 fps, a QVGA (320x240) depth camera which can record video at up to 60 fps. SoftKinetic is a short range camera which operates between 15 cm to 100 cm range from the device. It has a field of view of 74 by 58 by 87 (Horizontal, Vertical, Diagonal). The size of the camera makes it possible to be combined with another camera and still be able to be mounted on a HMD.

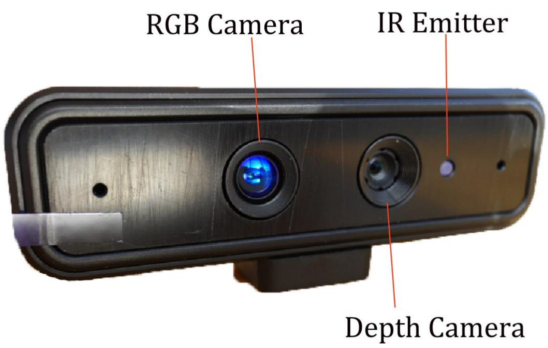


FIGURE 4.3: SoftKinetic, DS325

4.2.3 Hybrid Solution, Leap Motion plus SoftKinetic

To include real hand interaction with actual hand textures in real time. We need to create a hybrid solution. To create the hybrid solution, both cameras have to be combined on one hardware platform. This can be done using a SoftKinetic's 3D printable bracket which is available on their website[1]. SoftKinetic camera is screwed on the bracket and then mounted on top of Oculus DK2. The bracket attachment process can be seen in Fig. 4.4a. Similarly, a 3D printable mount [2] is also available for Leap Motion controller which can be glued on top of SoftKinetics bracket. The Leap Motion controller mount is shown in Fig. 4.4b, and the resulting hardware in Fig. 4.4c.

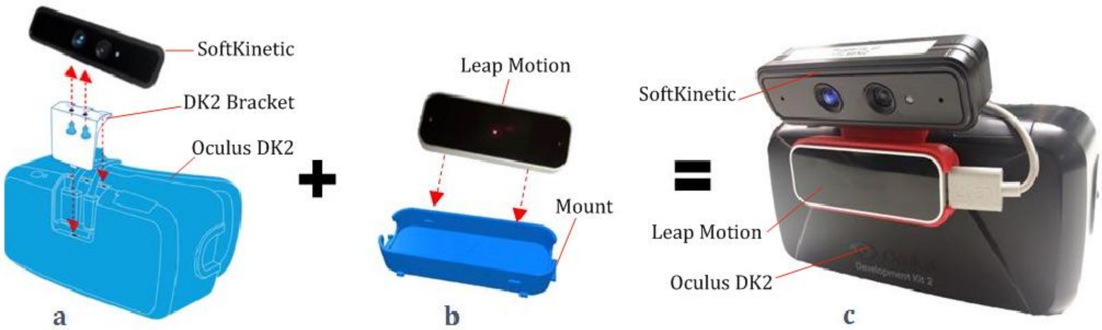


Figure 4.4: (a) SoftKinetic, DS325, attachment bracket (b) Leap Motion controller mount (c) Final hardware prototype

4.2.4 Head Mounted Display

Oculus Developer Kit 1 was the headset which started the third wave of virtual reality, so a lot of developers used Oculus Developer Kits to develop applications and try different hardware. Hence, it has more software and hardware platform compatibility than the other HMDs. Unity also has a built-in support for Oculus Developer Kit 2 (DK2). For our prototype, we used Oculus DK2 as it has a bracket available for SoftKinetic camera which makes it hardware compatible and Unity built-in support makes it software compatible.

[1] `http://www.reachvr.com/brackets`
[2] `http://www.thingiverse.com/thing:445866`

4.2.5 Desktop

A comfortable immersive virtual reality experience requires good computing power. The recommended computer hardware for Oculus is Intel i5-4590 equivalent or greater processor, 8GB RAM or more, a NVIDIA GTX970 or AMD 290 equivalent or greater graphics card. The computer used for building our prototype has Intel i7-6700, 3.4GHz processor, 16GB RAM, and a NVIDIA GTX980 graphics card.

4.3 Software

4.3.1 Unity3D Game Engine

Unity3D is a game engine and a cross-platform development software which allows developer to create graphics in both 2D and 3D. Due to its cross-platform nature, it supports easy integration with external hardwares such as depth and colour cameras, VR HMD (Oculus DK2, Samsung Gear VR). Unity also has VR API integrated within the platform which allows an easier development of a VR application. Unity is one of the first platforms which supported built-in VR development so it has been extensively used by the VR developer community, creating a lot of resources and tools developed in Unity. As Unity is a game engine, it already has physics engine and 3D graphics embedded which are required for creating immersive VR experiences. With all these capabilities, Unity was chosen as a development platform for our project. With its support for multiple software and hardware, it also served as a integration platform for our project.

4.3.2 DepthSense SDK

DepthSense SDK provides a software interface with Softkinectic, DS325 camera. Using the SDK, we can access RGB data, depth map, UV map and audio output from the camera. The SDK can be used to configure the colour data compression format to YUY2 or MJPEG. It can also change depth sensor capturing frame rate between 25 fps to 60 fps and colour camera between 25 fps to 30 fps [47]. Since the maximum colour camera frame rate is 30 fps which was used for our project. The SDK also decides the range of the camera whether close range or long range. For this project, a close range camera would be more applicable, so the camera was configured to close range settings.

In order to use Bai et al.'s algorithm [12], we needed the RGB data stream and the depth map stream and the UV map stream from the SoftKinetic camera. The data accessed from the camera using DepthSense SDK can be seen Fig. 4.5. For our application, we are using the Unity 5 game engine so the data streams need to interface with Unity 5. A dynamic link library (dll) of the algorithm was created to import the required data streams into Unity. To increase the efficiency of the process, a 64-bit version of the file was created and the 64-bit version of Unity 5 was used.

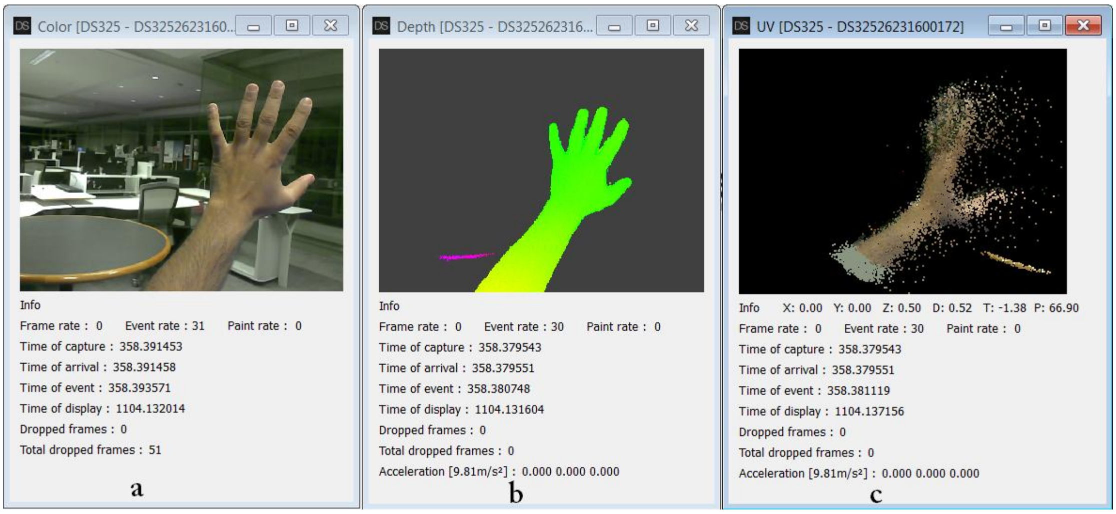


FIGURE 4.5: (a) RGB data (b) Depth map (c) UV Map

4.3.3 Leap Motion SDK

The Leap Motion controller is the hand tracker for our prototype, and the tracking information is accessed through the Leap Motion SDK. The SDK has two basic libraries which define the application programming interface (API) with the tracking data. One library is based in C++, and the other is in C. There are wrapper classes for these libraries for C#, Unity's scripting language. For Unity 5, there is a plugin that uses C# wrapper class to interface with the SDK and we used the plugin to access the hand tracking data.

Leap Motion's Unity plugin first transforms the tracking data from Leap Motion space to Unity space, then it applies the data to a virtual 3D hand model in Unity space. This acquisition and application of tracking data drives the hand model in Unity. The tracking data is reflected in the hand and finger position in the Unity. There are two different virtual hand models available in Leap Motion's Unity assets, graphics hand models and physics hand models. The graphics hand models are for showing the tracked hands without having any rigidbody properties, for example, collision or other physics based interactions. Physics hand models are invisible with the rigidbody properties, they are added along with the graphics hand models when user wants to interact in the virtual environment. These are few different graphics as well as physical hand models available in Leap Motion's Unity assets. Some of them can be seen in Fig. 4.6.

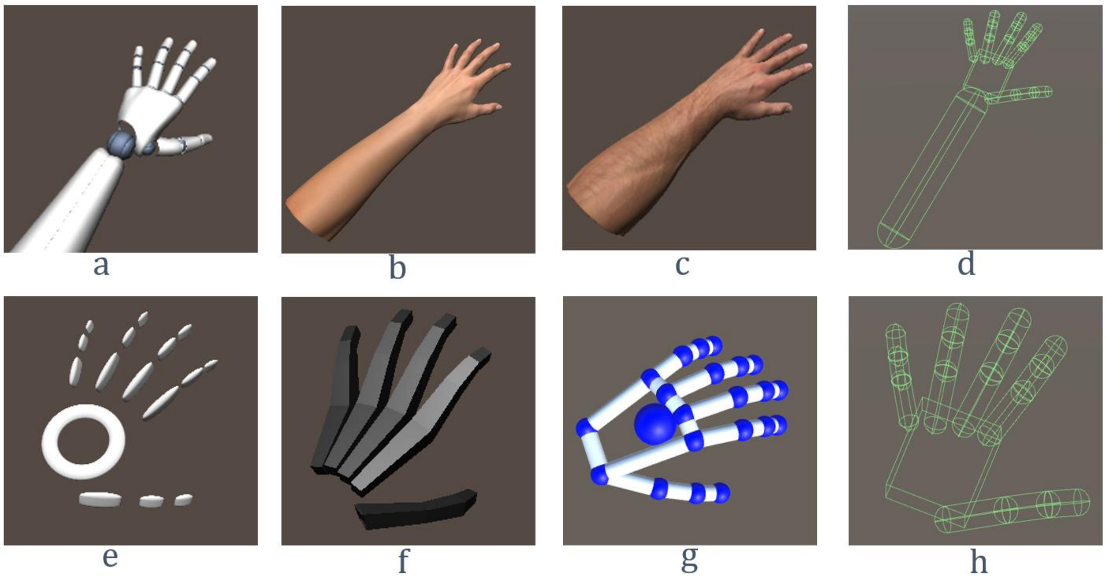


FIGURE 4.6: (a) Robot Hand (b) Feminine Hand (c) Masculine Hand (d) Full Physics Hand (e) Minimal Hand (f) Procedural Hands (g) Capsule Hand (h) Physics Hand

Leap Motion plugin requires the use of Leap Motion's camera asset, LMHeadMountedRig, for VR development. The LMHeadMountedRig matches the HMD movement within its tracking space, and also helps correctly place the hand models in the virtual space. In order to align the physical world with the virtual world, LMHeadMountedRig should be placed at a same location as its corresponding Leap Motion device location in the real world. The HMD position is the common frame of reference between the two worlds, and it helps with the alignment of two worlds, as illustrated in Fig. 4.7.

FIGURE 4.7: Leap Motion and HMD position in real and virtual world [13]

4.3.4 Hybrid Solution, Leap Motion plus SoftKinetic

With the two cameras combined on a single hardware, they are to be aligned in software. The software implementation involves capturing data from each camera by using their respective SDK, DepthSense SDK for SoftKinetic and Leap Motion SDK for Leap Motion controller. The captured streams from the two cameras are placed in Unity space to be processed and visualised. Once they are in the same 3D virtual space, the transformation between the two is computed to be synchronised. The transformation is based on a point cloud matching algorithm which uses singular value decomposition to align the cameras. With two cameras aligned together, we can see our real hand from SoftKinetic's RGBD camera and use Leap Motion's underlying physics-based hands to interact with the virtual environment. The process of importing the SDKs is explained below.

Software combination involved putting the data from the two cameras in one space which was achieved by using the two SDKs as explained above. After the data is in one space (Unity space) the transformation between the two data is computed using a point cloud matching algorithm [48]. The computed transformation is then applied on SoftKinetic's data to align with the Leap Motion data. Once they are aligned, we can show SoftKinetic's RGB data for visualisation and use Leap Motion's physics hand collider for interaction. Since the relative position between the two cameras remain fixed, we only have to compute the transform once and re-calibrate the cameras if there are changes.

The reason for using point cloud matching algorithm is that Bai's algorithm gives us the hand tracking data (finger tips and palm position) in the form of points in a 3D space, and these points can be matched with Leap Motion's tracking data. The points are in Unity's Softkinectic local space. In order to find the transform with the Leap Motion hand tracking data, we need to find the corresponding points in the Unity's Leap Motion local space. As the Leap Motion data is already in Unity, it can be accessed through a C# script. SoftKinectic's and Leap Motion's hand tracking data can be seen in Fig. 4.8. The tracking data from both cameras also contain the information about the right or left hand and the type of finger (thumb, index finger, little finger). By having this information, we know the correspondence between the two tracked data, which allows us to compute the transformation matrix between them.

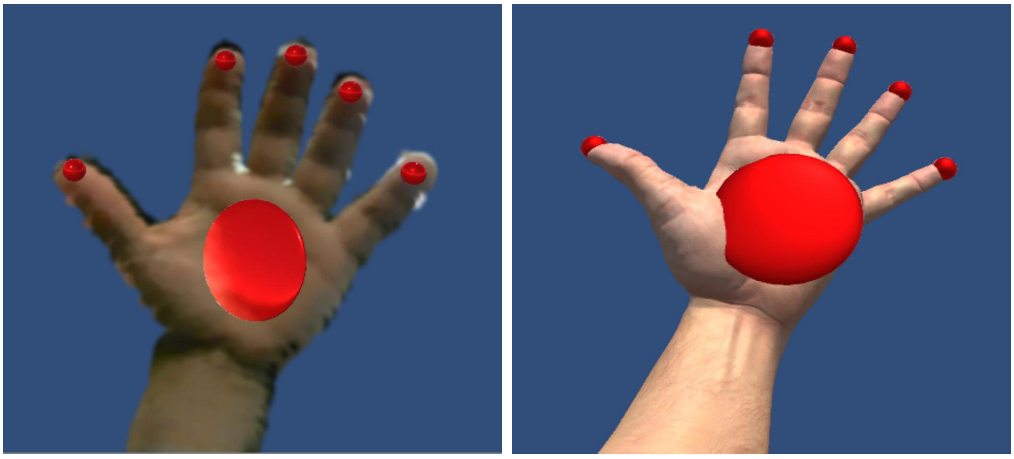


FIGURE 4.8: SoftKinectic(left) and Leap Motion(right) hand tracking data and corresponding points

The transformation matrix is computed using the SVD-based point cloud correspondence algorithm [48].The algorithm requires at least three corresponding points. For calibration, we considered one hand; therefore, we have six corresponding points (five fingers and one palm position). The six points can be seen in Fig. 4.8. The points from SoftKinetic are on the left and from Leap Motion on the right. We want to find the transformation matrix (rotation and translation matrix) that will transform six SoftKinetic points to their respective Leap Motion points and then apply this transformation matrix on the SoftKinetic data to align with the Leap Motion data.

Using the algorithm [48], the rotation and translation matrices are computed as following:

1. First, the centroids are calculated for both data sets.

$$SKPoint_i = \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} \qquad LMPoint_i = \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix}$$

$$centroid_{SK} = \frac{1}{N}\sum_{i=1}^{N} SKPoint_i \qquad centroid_{LM} = \frac{1}{N}\sum_{i=1}^{N} LMPoint_i$$

2. Then, the co-variance matrix, H.

$$H = \sum_{i=1}^{N} (SKPoint_i \; - \; centroid_{SK})(LMPoint_i \; - \; centroid_{LM})^T$$

3. The co-variance matrix, H, is factorised using singular value decomposition.

$$[U, S, V] = SVD(H)$$

4. From the factorised H matrix, rotation matrix,R, is computed by doing a dot product between V and transpose of U matrix which are obtained from SVD.

$$R = VU^T$$

5. The computed R matrix is the rotation matrix if the determinant of R is greater than zero. If it is less than zero, it is a reflection case. Then the third column of R matrix is multiplied by -1 which give us the rotation matrix. The pseudo-code:

$$if\ determinant(R)\ <\ 0$$

$$multiply\ 3rd\ column\ of\ R\ by\ -1$$

$$else$$

$$R\ remains\ the\ same$$

6. The translation matrix, t, was computed using the following equation:

$$t = -R * centroid_{SK} + centroid_{LM}$$

After obtaining the rotation and the translation matrices, they are applied on initial six points to compute Root Mean Square Error value, RMSE. The RMSE value helps us evaluate the accuracy of calculated rotation and translation matrices, and it is calculated using this equation:

$$RMSE = \sum_{i=1}^{N} ||R * SKPoint_i + t - LMPoint_i||^2$$

If the RMSE value is less than the threshold value, the optimal solution is found. If the value is greater than the threshold value same process is repeated from step 1 to 6 with different SoftKinect and Leap Motion data sets until the optimal solution is found. The whole process is illustrated in Fig 4.9.

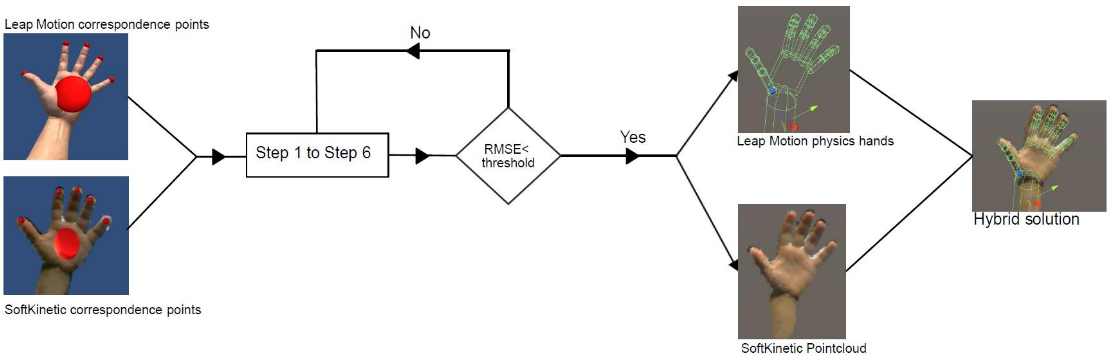


FIGURE 4.9: Flow chart illustrating point cloud alignment

With hands aligned, we can show user's real hands captured by SoftKinetic and use Leap Motion physics-based hands to interact with the the virtual world. This creates an augmented virtuality experience as user is able to view his/her real hands and interact with the surrounding virtual world. Now, with interactable real hands implemented, we need to combine them with a 360 spherical video player to create an interactive cinematic environment.

4.3.5 Projective Texture

At this point, the real hands are displayed as a point cloud in 3D space which can be discontinuous at places, as can be seen in Fig. 4.10. We tried to solve the problem by creating a projective texture using only the RGB information from the SoftKinetic camera and projecting it on to a pre-rendered Leap Motion hand models. The process is shown in Fig. 4.11.

FIGURE 4.10: Point cloud discontinuities

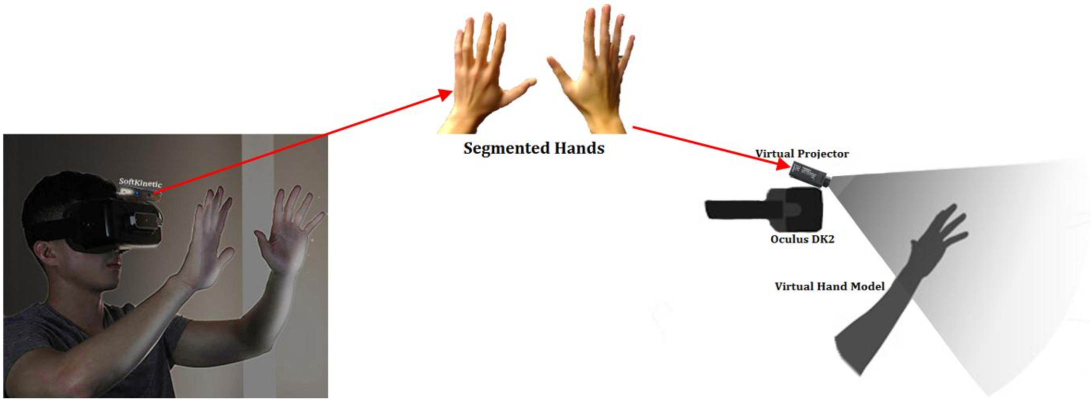


FIGURE 4.11: Projecting segmented RGB data on to a 3D hand model

However, the 3D geometry of the real hand does not match with the pre-rendered hand models so there are some parts of the RGB image that does not map on the hand models. This results in a mismatch between the projected texture and the 3D hand model which in turn makes some part of the hand model visible. The mismatch can be seen in Fig. 4.12. The hand models with projective texture were tried inside the lab, but the user experienced break in presence due to visible mismatch. Hence, the projective texture was not used for our prototype. At that point, we improved the real hands' point cloud representation by varying the point size and the other parameters of the particle system (Unity component to render point cloud). This improved the visual quality of the point cloud to an acceptable level for a user study.

FIGURE 4.12: Projective texture mismatch

4.3.6 Creating 360 Spherical Video Player

With hand representation finalised, we moved onto creating a 360 video player. Most of the 360 videos available online are equirectangular mapped videos which are spherical coordinate videos projected onto a planar coordinate. The video player assets available on Unity support 360 spherical videos playback as they are planar videos which can be mapped onto a sphere to create a 360 immersive video. However, the video player assets only support playing the video, a spherical environment still has to be created in Unity. To create the 360 spherical video player, we simply have to use Unity's sphere primitive and texture map the planar video sequences onto it. Since the sphere already has the vertices and texture coordinates, it makes it easier to map equirectangular videos. However, mapping is done only on the outside of the sphere so the user viewing from the centre from the sphere would not be able to see the mapped video. There is a way to fix it by culling the front faces and showing the back faces of each polygon being rendered. This was done by a Unity shader that shows the video from an inside perspective. The process is illustrated in Fig. 4.13. For the Unity video player asset, we tried three different assets: AVPro Video, Easy Movie Texture and OpenVR dev kit for SteamVR.

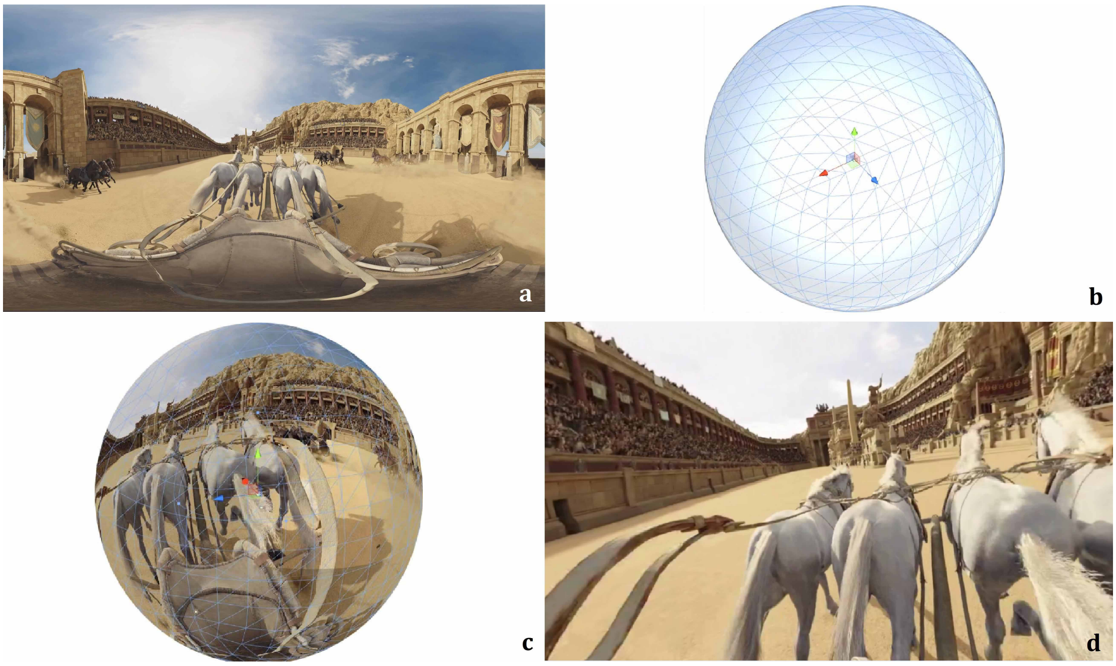


Figure 4.13: 360 VR player video mapping, a) Equirectangular mapped video b) Mapping Sphere c) Video player with inside mapping d) Viewer's perspective

The AVPro is a paid asset but a trial version can be used with watermarks. We tried it and found the watermark to be distracting for an immersive experience. So we looked into the OpenVR dev kit, a free asset which uses SteamVR to play the video. OpenVR plays the 360 video without any watermark but it requires Oculus DK2 to be constantly facing the head tracking camera. If it is not facing the camera, the virtual camera (the viewer point of view) is initialised out of the sphere which causes the display screen to go blank. Therefore, we looked into another video player, Easy Movie Texture. Easy Movie Texture did not have any watermark and was streaming video output without going blank when Oculus DK2 is not facing the camera, so we decided to use Easy Movie Texture for our prototype. With the 360 video player finalised, our software to play the 360 video and key in users hands is finalised. Now, we could focus on the content (the 360 movie, interactable objects, and the hand gestures for creating the immersive interactive experience).

4.4 Content

4.4.1 360 Videos for Interactive Immersive Experience

The 360 video is the main ingredient of our project, as it contains the narrative of immersive experience and also specifies the user role in the story. For our project, the user has a participant passive role in the video that is the user can view as well as interact with the environment, but the interaction will not change the outcome or narrative of the video; it will simply add flavour to the experience[10]. For an observant active role to work, the user has to make sense in the environment, and the interaction has to be plausible enough to keep the presence intact. The background of the video realistic or animated was also considered as the user hands were going to be shown in the environment so they should make sense with the environment. Since we were planning to use realistic computer generated hands and the real hands, we considered both types of video, realistic (real world imagery) as well as animated (cartoony imagery). With these guidelines in mind, we found these two videos which could possibly be used for our prototype, "Clash of Clans" [3], "Wild Dolphins" [4].

Creating interactive experience seemed more plausible with these videos, as the observant (shooting camera) is playing a role in the environment. For the "Wild Dolphins" video, screenshot shown in Fig. 4.14, we thought of adding a swimming gesture (bringing both hands together and then move them apart). The user swims backward and forward doing this gesture. Another interaction, we thought was creating virtual bubbles beneath the user, and the user interacts with the bubbles using hands. However, creating immersion with the realistic underwater scene is less plausible as there is no water surrounding the user so we looked into "Clash of Clans" video

[3] `https://www.youtube.com/watch?v=wczdECcwRw0`
[4] `https://www.youtube.com/watch?v=BbT_e8lWWdo`

Figure 4.14: "Wild Dolphins" screenshot

Next was the “Clash of Clans” video, we created a simple interaction for the video and tried it out. The video is a game scene where there is a battle going on so we placed bombs infront of the user which could be picked and thrown at the characters coming toward the user. The screenshots of the interaction can be seen in Fig. 4.15. However, we decided not to use the video as it is an animated video. User expectation for interaction would be more with computer generated videos than the realistic video because the recorded graphics seem interactable but they are not which will be frustrating and can potentially break presence. At this point we also decided not to use animated videos for our prototype.

Figure 4.15: "Clash of Clans" interaction

Since both videos did not reach our criteria, we started looking for other 360 videos. We took out animated videos from our search and looked for realistic video. We tried hot air balloon 360 videos. The videos seemed plausible for creating in-air interaction such as throwing objects towards the ground or interacting with flying objects. With this new idea we found these two videos which are suitable for creating in-air interaction, "Ballooning Bad BirnBach" [5],"Take a Journey on a Hot Air Balloon" [6]. We finalised "Ballooning Bad BirnBach" for our prototype (Fig. 4.16), as it had more potential to add interaction. After deciding on the 360 video, we moved on to creating 3D interactable objects and interaction with them using hand gestures.

FIGURE 4.16: "Ballooning Bad BirnBach" screenshot

4.4.2 Hand Gestures and 3D Interactable Objects

To create an interactive immersive experience in VR, we need to add 3D interactable objects. These objects should be related to the movie being played on the 360 video player to keep the illusion intact. The 3D objects can only be introduced in the space between viewer and the spherical movie screen, the interaction space is shown in Fig. 4.17. The 3D interactable objects should add-on to the narrative or relate to the narrative, in order to keep it immersive. They should follow the movie's norm to make them more plausible in the environment. In our case, the movie was taken on a hot air balloon, and it is a calm hot air balloon ride with a simple narrative.

[5] `http://www.vrideo.com/embed/bhmZwIvI`
[6] `https://www.youtube.com/watch?v=0Q-UOM4FjdM`

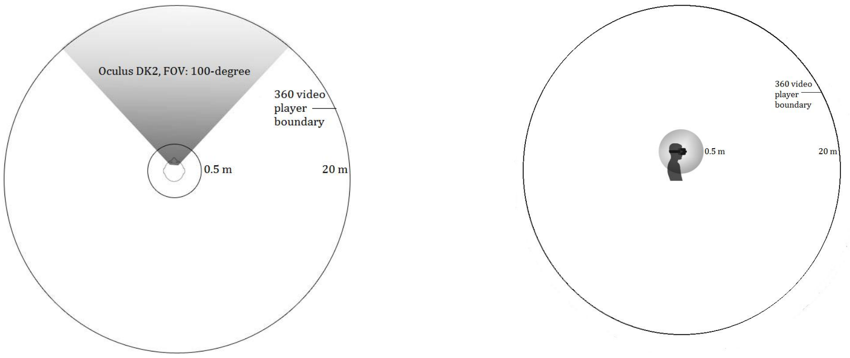


FIGURE 4.17: Interaction space, top view on the left and side view on the right

We thought of creating in-air interaction, such as dropping objects towards the ground or interacting with flying objects. We created a flying Aladdin's gift rug. The user pick-up gifts on the rug and throw on the houses below. User can also create flowers using hand gesture. For gifts, the user directly uses his/her hand (direct hand manipulation) to pick-up gift and throw it down towards the houses. The interaction is illustrated in Fig. 4.18. The thrown gifts can be regenerated by doing a left hand open gesture shown in Fig. 4.19. With right hand open gesture, user can create flowers falling from the sky, as can be seen in Fig. 4.19. However, after testing we decided not to use this interaction for our prototype as the video was not Christmas themed, and it was not making sense as a whole experience.

FIGURE 4.18: Christmas gifts interaction

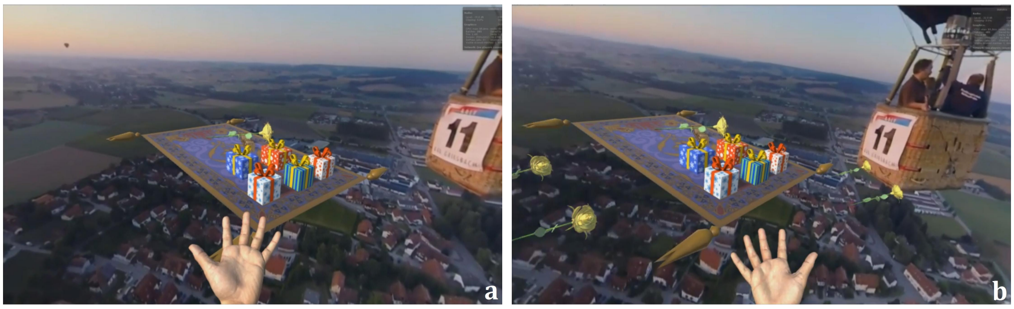


FIGURE 4.19: Hand gesture-based interaction, a) Left hand b) Right hand

The first in-air interaction did not work so we brainstormed for more ideas, and thought of interaction with flying birds. To add flying birds, we need a rigged 3D bird model which can fly and perform basic animations. We found a bird asset [7] in Unity store and used it to create basic flying and gliding within the spherical video space. Then we added in a feeding interaction using a simple hand gesture. The hand gesture was simple, open and close the hands facing upward, shown in Fig. 4.20. With this gesture, food was created simply above the hand. The user grabs and throws the food towards the bird, and the birds try to catch it by gliding downward towards the food. The user also pours food from one hand to the other, refer to Fig. 4.20. The feeding interaction is explained in Fig. 4.21. This interaction is more plausible and birds are making sense in the whole environment.

FIGURE 4.20: Hand gesture to create food and interaction with other hand

[7] https://www.assetstore.unity3d.com/en/#!/content/36951

(a) Birds flying on right side of user

(b) Birds flying on left side of user

(c) Hand gesture to create food

(d) Attracting birds with food

(e) Birds flying down to collect food

(f) Birds flying back with food

(g) Flying back to original position

(h) Restart normal flying pattern

Figure 4.21: Birds and the feeding interaction

4.4.3 Trial Run

After finalising the prototype, we tested it with three HIT Lab students to obtain feedback. During the trial run, it was found that the current hand gesture was distracting to the user. When users tried to grab the food in their fist, it was slipping away. This was causing users to focus on simply grabbing the food. The food was slipping away because Leap Motion physics hands are made of cylinders (fingers and thumb) and a cuboid (palm). The collision between the spherical food and cylindrical fingers was causing it to slip out of the hands. We came up with a solution to solve this problem and changed the hand gesture to having the hands facing downwards. The food would then be generated from the hand facing downwards, and with opening and closing hand gesture the food falls down. The old and new hand gesture is shown in Fig. 4.22.

(a) Old gesture hand close

(b) Old gesture hand open

(c) New gesture hand close facing downwards

(d) New gesture hand open facing downwards

FIGURE 4.22: Old and new hand gesture to create food

4.5 Summary

The development of the prototype was described in this chapter. First, we combined the Leap Motion and SoftKinetic camera on a hardware platform, then on the software platform. We also created the 360 video player with Unity's sphere primitive and the Easy Movie texture plugin. After finalising the underlying software and hardware, we developed the content for the prototype. We selected a 360 video for creating a plausible interaction. Then we designed the interaction (interactable object and hand gesture) for the 360 video. And finally, we did a trial run for the prototype and updated the hand gesture for the interaction.

Chapter 5

User Evaluation

This chapter describes an experiment which uses the created prototype to investigate the effect of interaction and a user's hands appearance on user's presence and embodiment in a cinematic 360 VR scene. The chapter is divided into three sections: evaluation purpose, hypothesis and the user experiment design.

5.1 Evaluation Purpose

The purpose of the evaluation is to understand whether the interaction and a hand's appearance affect the user's presence and embodiment in a cinematic 360 VR scene. We are particularly interested in finding out whether adding interaction improves presence and embodiment in the scene. We are also interested in studying the effect of augmenting virtual hands with real hands on user's presence and embodiment.

5.2 Hypotheses

We have six hypotheses for the experiment which are as follows:

- Hypothesis 1 (H1): There is a significant difference in the sense of presence with the hand gesture-based interaction and without the hand gesture-based interaction, when experiencing an interactive 360 movie.

- Hypothesis 2 (H2): There is a significant difference in the sense of presence between viewing one's own hands and realistic computer generated hands, when experiencing an interactive 360 movie.

- Hypothesis 3 (H3): There is a significant difference in the sense of agency with the hand gesture-based interaction and without the hand gesture-based interaction, when experiencing an interactive 360 movie.

- Hypothesis 4 (H4): There is a significant difference in the sense of agency between viewing one's own hands and realistic computer generated hands, when experiencing an interactive 360 movie.

- Hypothesis 5 (H5): There is a significant difference in the sense of ownership with the hand gesture-based interaction and without the hand gesture-based interaction, when experiencing an interactive 360 movie.

- Hypothesis 6 (H6): There is a significant difference in the sense of ownership between viewing one's own hands and realistic computer generated hands, when experiencing an interactive 360 movie.

5.3 User Experiment Design

5.3.1 Factorial Design

There are two independent variables for the experiment and each has two levels so four conditions in total. The factorial design for the experiment is 2 by 2, as shown in Table 5.1.

	Interaction with the movie	No Interaction with the movie
Real Hands	A	B
CG Hands	C	D

Table 5.1: Factorial design

5.3.2 Balance Latin Square

The experiment is a within-subject design for both factors so each participant will try all four conditions. As the experiment features within-subject design, we used balance latin square design to counter-balance the order effect. There are four conditions for the experiment; therefore, 4 by 4 balance latin square design was used, which can be seen in Table 5.2.

A	B	C	D
B	D	A	C
D	C	B	A
C	A	D	B

TABLE 5.2: 4 by 4 Balance Latin Square Design

where	A	=	Real Hands	+	Interaction
	B	=	Real Hands	+	No Interaction
	C	=	CG Hands	+	Interaction
	D	=	CG Hands	+	No Interaction

5.3.3 Experimental Setup

The study was carried out at the Student Lab in HIT Lab NZ. The setup can be seen in Fig. 5.1.

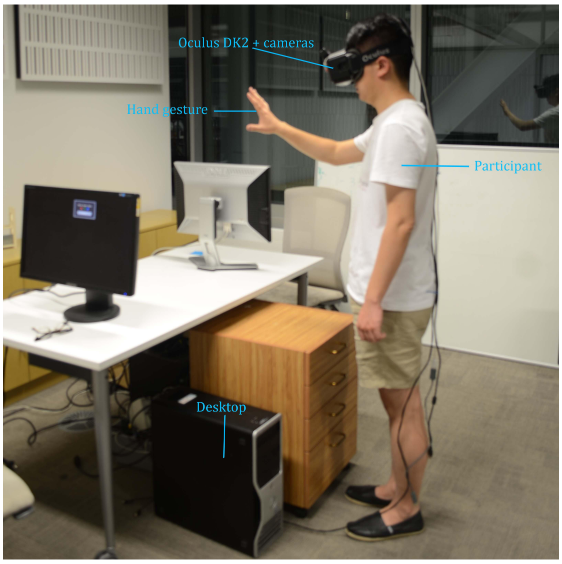


FIGURE 5.1: Experimental setup

The material used for the experiment is the same as the one described in chapter 4, an Oculus DK2, a Leap Motion controller, a SoftKinetic DS325 camera and a desktop. Because we are doing a within-subject study, we do not want to exhaust our users so the 360 video in Chapter 4.3.1 ("Ballooning Bad BirnBach") is trimmed down to 2 minutes. The two minutes give ample time to observe the surroundings and to perform interaction with the virtual environment. To shorten the video, we selected the video sequence which preserves continuity in the scene and avoids any jumps in audio or video sequence. For the CG hands, we want users to select a hand model which most closely resembles their own hands. There are six CG hand models available for the user experiment, three for each gender and depicting three different skin tones: white, brown and black. The six CG hand models can be seen in Fig. 5.2.

FIGURE 5.2: The six CG Hand Models

5.3.4 Procedure

The experiment followed the procedure describe below. Overall the experiment took 40 minutes. 1. At the beginning of the experiment participants are given a general overview plus explanation of the project. This is followed by an information sheet and the consent form to read and sign. The information sheet and the consent form are provided in Appendix A.

2. They are then given a pre-experiment questionnaire to answer. The pre-experiment questionnaire collects demographic information and can be found in Appendix B.

3. The pre-experiment questionnaire is followed by a detailed briefing on the experiment which includes a description of the experimental tasks, the gesture for interaction and the experimental process.

4. Participants are then provided with the prototype, an Oculus DK2 with two cameras (Leap Motion plus SoftKinetic), to go through a demo experience. In the demo experience, they choose the CG hands which they are going to use for interactions and also practice the feeding hand gesture. The participants are given sufficient time to practice the gesture. Once they are comfortable with it, they are shown their real hands to minimise any novelty effect when seeing their real hands. They also practise the same hand gesture with their real hands.

5. At this point, participants are ready to start the actual experiment. They undergo the experiment's four conditions in a sequence defined by the balance latin square chart. At the end of each condition, the participants answer a per-condition questionnaire to rate their experience.

6. After completing the four conditions, they answer a post-experiment questionnaire to give feedback on their overall experience.

7. Lastly, there is a debriefing session to clarify any issues in the questionnaire responses and a short interview to discuss their interactive and non-interactive experiences with different hand representations.

5.3.5 Experimental Task

The experiment is within-subject design as explained above, so each participant goes through all the four conditions, A, B, C and D, as shown in Fig. 5.3. The tasks for each condition are as follows.

Condition A: The participant is shown the "Ballooning Bad BirnBach" 360 video and the real hands. They can use the hands for interaction in the virtual environment. The interaction in our experiment is feeding the birds using hand gesture. After the two minutes of video, they answers the per-condition questionnaire.

Condition B: In condition B, the participant watches the "Ballooning Bad BirnBach" 360 video and sees their real hands. However, for condition B, there is no interaction. After the video, they answers the per-condition questionnaire.

Condition C: In condition C, the participant is shown the CG hands which they chose during the demo. Similar to condition A, they watches the 360 video and interacts with the virtual environment using hands; however, in Condition C, s/he uses the CG hands instead of real hands. The video is followed by a questionnaire.

Condition D: For condition D, the participant watches the 360 video with CG hands. He can see their CG hands, but there is no interaction with the virtual environment. The participant answers the questionnaire after the video.

FIGURE 5.3: Condition A(top left), B(top right), C(bottom left) and D(bottom right)

5.3.6 Measures

There are mainly two measures recorded for the experiment, sense of presence and sense of embodiment.

a) Sense of Presence

The first measure we are interested to study is sense of presence. It is a psychological state where a person experiencing a virtual environment has a feeling of being there in a virtual environment. Sense of presence can be measured in different ways, for example measuring brain activity, physiological measures or conventional questionnaire [49]. Most of these measure are specific to a certain application. We choose the igroup presence questionnaire (IPQ) [1] as it measures the presence components (general presence, spatial presence, realism and involvement) we are interested in and has been widely used in previous research. IPQ is composed of 14 items, which are rated on seven-point Likert scale [32]. These 14 items are further divided into three sub-scales and one general item. The three sub-scales are highlighted below.

- Spatial Presence: the sense of being physically present in the virtual environment
- Involvement: measuring the attention devoted to the virtual environment and the involvement experienced
- Experienced Realism: measuring the subjective experience of realism in the virtual environment
- General Presence: assessing the general "sense of being there"

These three scales are independent of each other. The fourth item, the general presence has an effect on all three sub-scales specially the spatial presence [50],[51],[52]. For our experiment, we used IPQ to measure presence for each experimental condition. The per-condition questionnaire is provided in Appendix B.

[1] IPQ `http://www.igroup.org/pq/ipq/items.php`

b) Sense of Embodiment

The second measure we studied is the sense of embodiment. We focused on two dimensions of the embodiment, the sense of agency and the sense of ownership. Sense of agency is the feeling of being in control of the virtual hands or the real hands i.e. the movement or action of hands in the real world translates into the virtual world. Sense of ownership is the feeling that the virtual representation of hands is one's own body. As sense of embodiment is a subjective feeling, also a questionnaire was used to measure it. The questionnaire from Argelaguet et al. work was adopted[15]. The questionnaire is part of per-condition questionnaire and can be found in Appendix B.

5.3.7 Pilot study

We conducted an initial pilot study on four participants, two males and two females. The pilot study tested all the conditions with the questionnaire. It was a full-on study in which participants went through same experience as the actual user experiment. The pilot study was carried out to detect any issue with the system or procedure for the experiment. The data collected from the pilot study was not included in the final data. During the pilot study, we found a software bug in a bird animation which was fixed for the user experiment so it further improved our system for user study.

Chapter 6

Results

This chapter presents the results obtained from the user study. The data collected from each participant consists of a pre-experiment questionnaire, four per-condition questionnaires and a post-experiment questionnaire. The pre-experiment questionnaire collected participant's demographic information, per-condition questionnaire gathered quantitative data for that specific condition and post-experiment questionnaire collected quantitative as well as qualitative data on participant's overall experience.

For presence, the overall IPQ data showed no significant difference between having hand gesture-based interaction and no interaction; however, there were results with significant difference in IPQ's sub-scales (realism, involvement) and a general item (general presence). For the sub-scales with significant difference, we found hand gesture-based interaction data had significant difference for general presence and involvement, and hand appearance data had significant difference for realism and involvement. For embodiment, we found significant difference in both agency and ownership when there is a hand gesture-based interaction. The hand appearance data had significant difference for ownership but no significant difference for agency.

6.1 Demographics

For the user experiment, 32 university students were recruited after receiving approval from the ethics committee (Appendix A). Among the 32 participants, 16 (50%) were males and 16 (50%) were females. Participants age varied between 19 to 42 years, with a mean of 26.97 and a standard deviation of 6.32. Out of 32 participants, 10 have never used HMD before, 11 have used it few times, and 6 are more frequent users. For CG hands, participants were given a choice between three different hand models with three varying skin colour, black, brown and white for each gender. This gave an option to choose equal number of males and females for the experiment which closely reflects the general population of the sample. Hence, the results of the experiment are more likely to be generalisable.

6.2 Quantitative Measures

For our experiment, we have two independent variables (IVs) each with two levels. The dependent variables are participants' response to presence and embodiment questions from a questionnaire. The questions are answered on a seven point Likert scale. As it is a Likert scale rating so the data collected is an ordinal data and a non-parametric test should be used to interpret it. With IVs in the experiment, we have to look for two main effects and one interaction effect between IVs. Common non-parametric tests (e.g. Kruskal-Wallis, Friedman) cannot be used to study the interaction effect. Hence, Aligned Rank Transform (ART) [53] was used to transform ordinal data, and two-way repeated measures analysis of variance (ANOVA) was applied on the transformed data. With two-way repeated measures ANOVA, we can study both main effects as well as interaction effect. This was done for both presence and embodiment data. Presence and embodiment's two components: agency and ownership, were analysed separately.

6.2.1 Igroup Presence Questionnaire, IPQ

The igroup presence questionnaire is a scale to measure the sense of presence in a virtual environment. The IPQ consists of 14 Likert-scale items which are rated on a scale of 1 to 7. IPQ further consists of four sub-scales, and for our study we also analysed results for each component.

For the data from 32 participants, we applied ART on raw data followed by a two-way repeated measure ANOVA. We found no significance difference in presence felt by users between having interaction or no interaction ($F(1,31) = 0.530$, $p = 0.472$), and the type of hands, real hands or CG hands ($F(1,31) = 2.221$, $p = 0.146$). There was also no interaction effect between the two IVs (F(1,31) = 0.027 , p = 0.870). The results are summarised in Table 6.1, and the box-plot for overall IPQ score is shown in Fig. 6.1.

IV's main effect and interaction effect	F(1,31)	Sig.
Interaction or No Interaction	0.530	0.472
Real Hands or CG Hands	2.221	0.146
Interaction or No Interaction * Real Hands or CG Hands	0.027	0.870

Table 6.1: Overall IPQ, two-way repeated measure ANOVA SPSS results

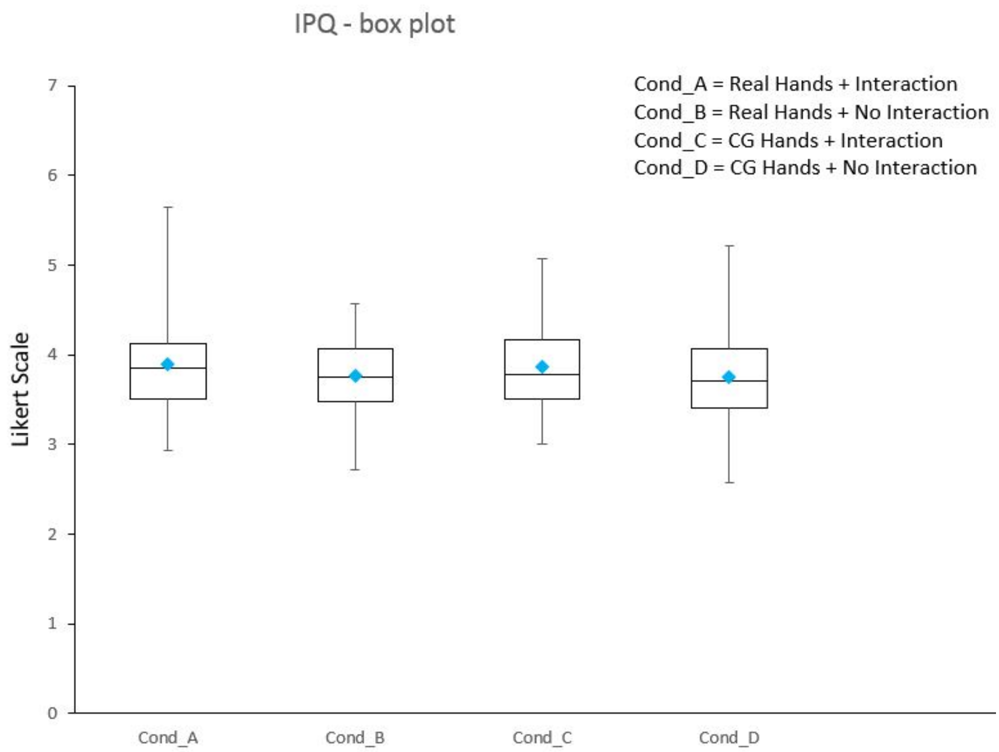


Figure 6.1: Box-plot for overall IPQ

We further analysed the four individual scales of the IPQ (General Presence, Spatial Presence, Involvement and Experienced Realism) and the results are as follow.

a) General Presence, GP

For general presence, using ART and two-way repeated measure ANOVA, we found significant difference between having interaction or no interaction (F(1,31) = 6.462, p = 0.016). However, there is no significance difference found for the type of hands, real hands or CG hands (F(1,31) = 0.021, p = 0.887), and also no interaction effect between IVs (F(1,31) = 0.000, p = 0.993). Results of the two-way repeated measure ANOVA are summarised in Table 6.2, and the box-plot for general presence is shown in Fig. 6.2.

IV's main effect and interaction effect	F(1,31)	Sig.
Interaction or No Interaction	6.462	0.016
Real Hands or CG Hands	0.021	0.887
Interaction or No Interaction * Real Hands or CG Hands	0.000	0.993

TABLE 6.2: General Presence, two-way repeated measure ANOVA SPSS results

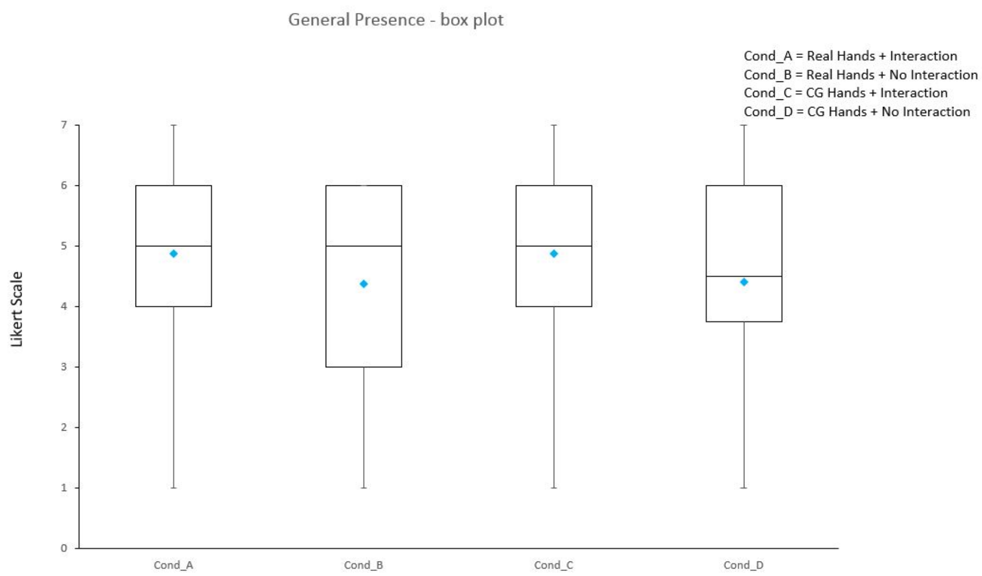


FIGURE 6.2: Box-plot for general presence

b) Spatial Presence

For spatial presence, after applying ART and two-way repeated measure ANOVA, we did not find any significant difference for the two main effects, interaction or no interaction (F(1,31) = 0.978, p = 0.330) and type of hands (F(1,31) = 0.003, p = 0.954), and the one interaction effect (F(1,31) = 0.011, p = 0.918). Results of the two-way repeated measure ANOVA are summarised in Table 6.3, and the box-plot is shown in Fig. 6.3.

IV's main effect and interaction effect	F(1,31)	Sig.
Interaction or No Interaction	0.978	0.330
Real Hands or CG Hands	0.003	0.954
Interaction or No Interaction * Real Hands or CG Hands	0.011	0.918

TABLE 6.3: Spatial Presence, two-way repeated measure ANOVA SPSS results

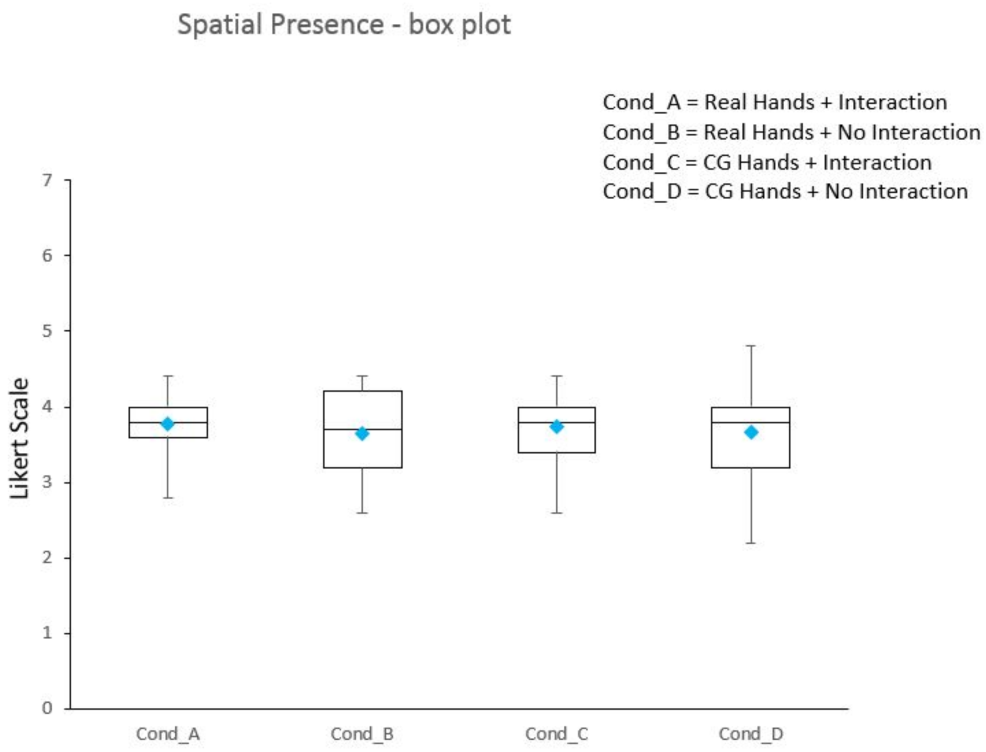


FIGURE 6.3: Box-plot for spatial presence

c) Involvement

Similarly, for involvement, we applied ART on raw data followed by a two-way repeated measure ANOVA. We found the significant difference for the two main effects, interaction or no interaction ($F(1,31) = 23.390$, $p = 0.000$) and type of hands ($F(1,31) = 9.373$, $p = 0.005$), as well as the interaction effect ($F(1,31) = 12.852$, $p = 0.001$). The results of two-way repeated measure ANOVA are summarised in Table 6.4, and the box-plot is shown in Fig. 6.4. Since there is a significant interaction effect between IVs, involvement line graph is also drawn in Fig. 6.5 to further understand the interaction effect.

IV's main effect and interaction effect	F(1,31)	Sig.
Interaction or No Interaction	23.390	0.000
Real Hands or CG Hands	9.373	0.005
Interaction or No Interaction * Real Hands or CG Hands	12.852	0.001

Table 6.4: Involvement, two-way repeated measure ANOVA SPSS results

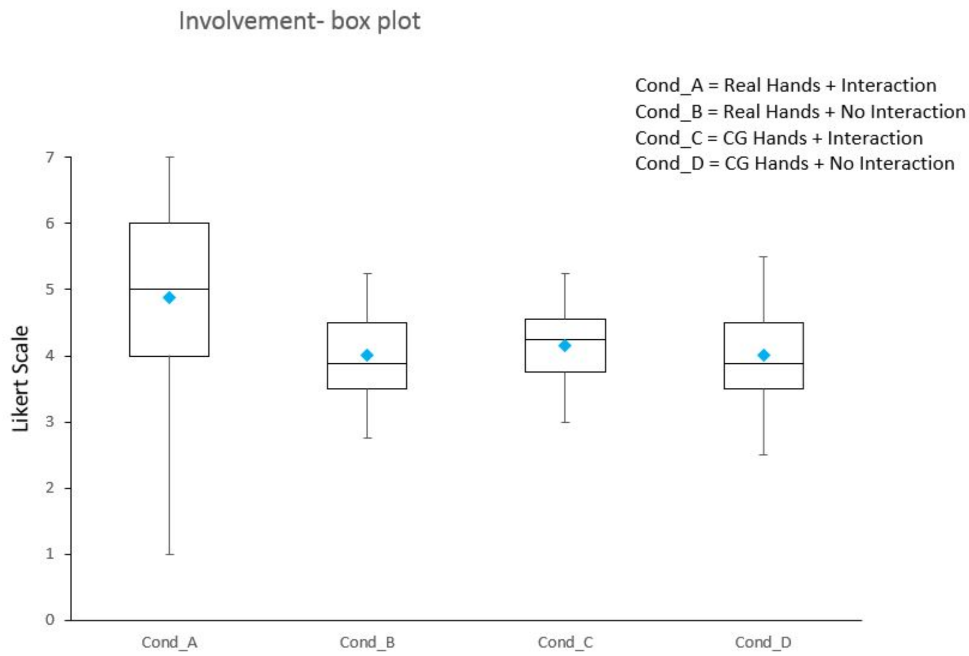


Figure 6.4: Box-plot for involvement

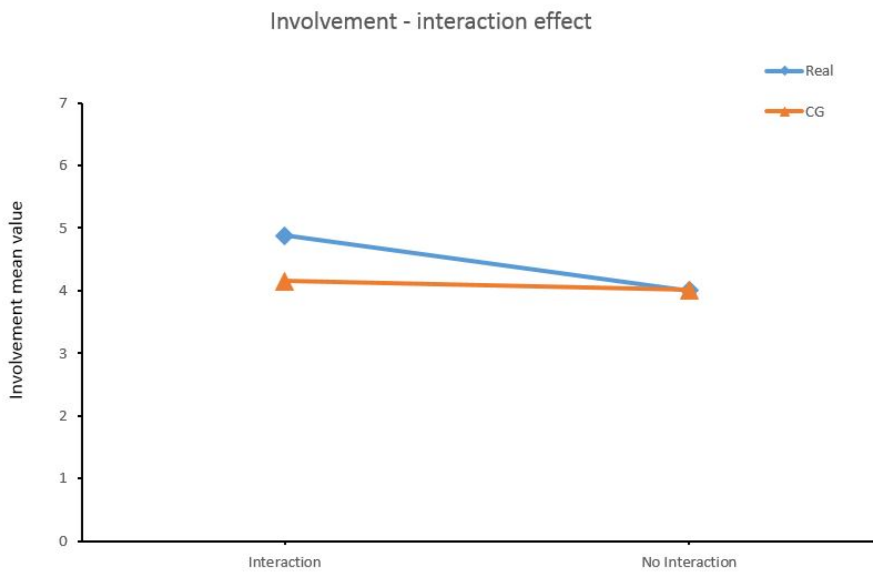


Figure 6.5: Line graph for involvement

d) Experienced Realism

For experienced realism, after applying ART and two-way repeated measure ANOVA, we found significant difference for one main effect i.e. for type of hands, real hands or CG hands ($F(1,31) = 4.209$, $p = 0.049$). We did not find significant difference for having interaction or no interaction ($F(1,31) = 0.338$, $p = 0.565$) and also no interaction effect between the factors ($F(1,31) = 0.087$, $p = 0.770$). The results of two-way repeated measure ANOVA are summarised in Table 6.5, and the box-plot is shown in Fig. 6.6.

IV's main effect and interaction effect	F(1,31)	Sig.
Interaction or No Interaction	0.338	0.565
Real Hands or CG Hands	4.209	0.049
Interaction or No Interaction * Real Hands or CG Hands	0.087	0.770

TABLE 6.5: Experienced Realism, two-way repeated measure ANOVA SPSS results

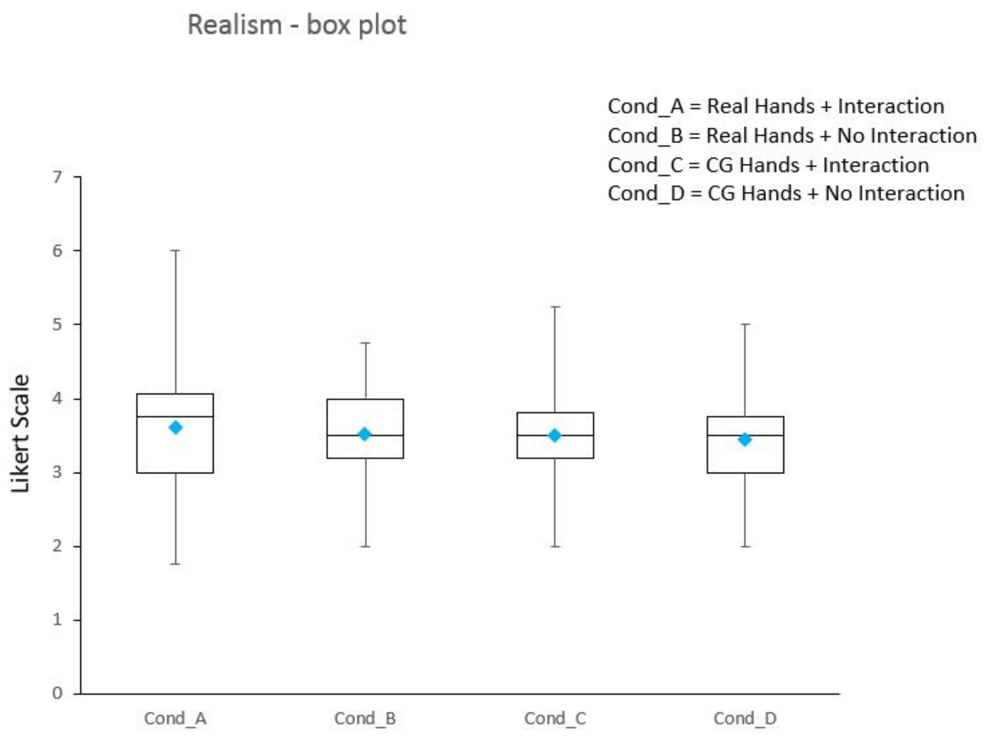


FIGURE 6.6: Box-plot for experienced realism

6.2.2 Embodiment Questionnaire

Embodiment was measured by evaluating user's sense of agency and sense of ownership with a questionnaire. The questionnaire is adopted from Argelaguet et al.[15] work, and both agency and ownership were measured separately. The results are as follow.

a) Agency

For agency, after applying ART and two-way repeated measure ANOVA, we found significant difference between having interaction or no interaction (F(1,31) = 12.256 , p = 0.001), and no significant difference for the type of hands (F(1,31) = 0.522 , p = 0.476). There is also no interaction effect between the two IVs (F(1,31) = 0.006 , p = 0.938). The results of two-way repeated measure ANOVA are summarised in Table 6.6, and the box-plot is shown in Fig. 6.7.

IV's main effect and interaction effect	F(1,31)	Sig.
Interaction or No Interaction	12.256	0.001
Real Hands or CG Hands	0.522	0.476
Interaction or No Interaction * Real Hands or CG Hands	0.006	0.938

Table 6.6: Agency, two-way repeated measure ANOVA SPSS results

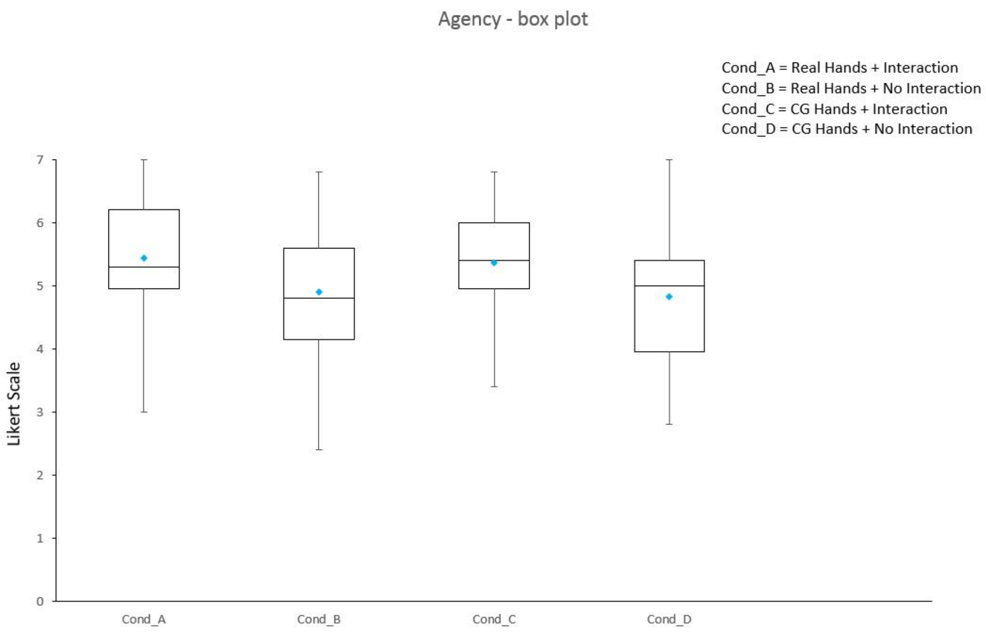


Figure 6.7: Box-plot for agency

b) Ownership

For ownership, using ART and two-way repeated measure ANOVA, we found significant difference for both main effects, interaction or no interaction ($F(1,31) = 9.290$, $p = 0.005$), and type of hands ($F(1,31) = 11.255$, $p = 0.002$). There was no interaction effect found between the two IVs ($F(1,31) = 0.124$, $p = 0.727$). The results of two-way repeated measure ANOVA are summarised in Table 6.7, and the box-plot is shown in Fig. 6.8.

IV's main effect and interaction effect	F(1,31)	Sig.
Interaction or No Interaction	9.290	0.005
Real Hands or CG Hands	11.255	0.002
Interaction or No Interaction * Real Hands or CG Hands	0.124	0.727

Table 6.7: Ownership, two-way repeated measure ANOVA SPSS results

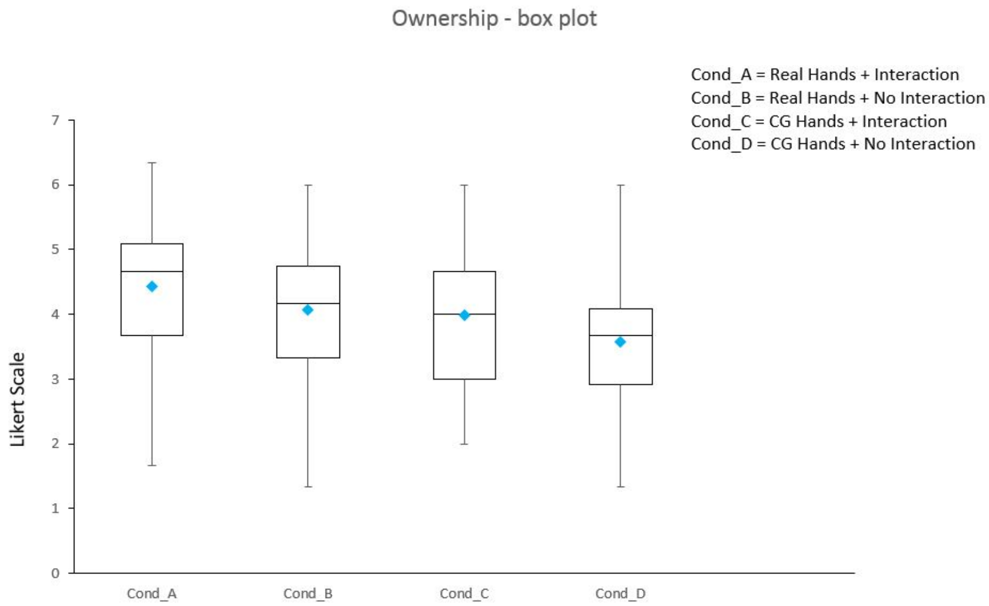


Figure 6.8: Box-plot for ownership

6.2.3 Post-experiment Questionnaire

After completing the four conditions, users were given a post-experiment questionnaire for their feedback on overall experience. We asked the users to rank the four conditions from 1:best to 4:worst. The bar chart for the ranking is shown in Fig. 6.9. We also performed a Friedman test on ranking data to find a significant difference, the test results are shown in Fig. 6.10. The Friedman test showed a significant difference in the rankings (χ^2 (3) = 48.0, p = 0.000), so we did a post hoc analysis on each condition pair using Wilcoxon signed-rank tests with a Bonferroni correction applied. There are four conditions so six pairs($^4C_2 = 6$) in total resulting in significance level at $p < 0.00833$ (0.05/6)). With the Wilcoxon signed-rank post hoc tests, we found significant difference for four pairs, Condition A and Condition B (Z = -4.580, p = 0.000), Condition A and Condition D (Z = -2.868, p = 0.004), Condition B and Condition C (Z = -4.614, p = 0.000), Condition C and Condition D (Z = -4.580, p = 0.000), and no significant difference for, Condition A and Condition C (Z = -2.183, p = 0.029), Condition B and Condition D (Z = -2.868, p = 0.046). The results of Wilcoxon signed-rank post hoc test are shown in Fig. 6.11

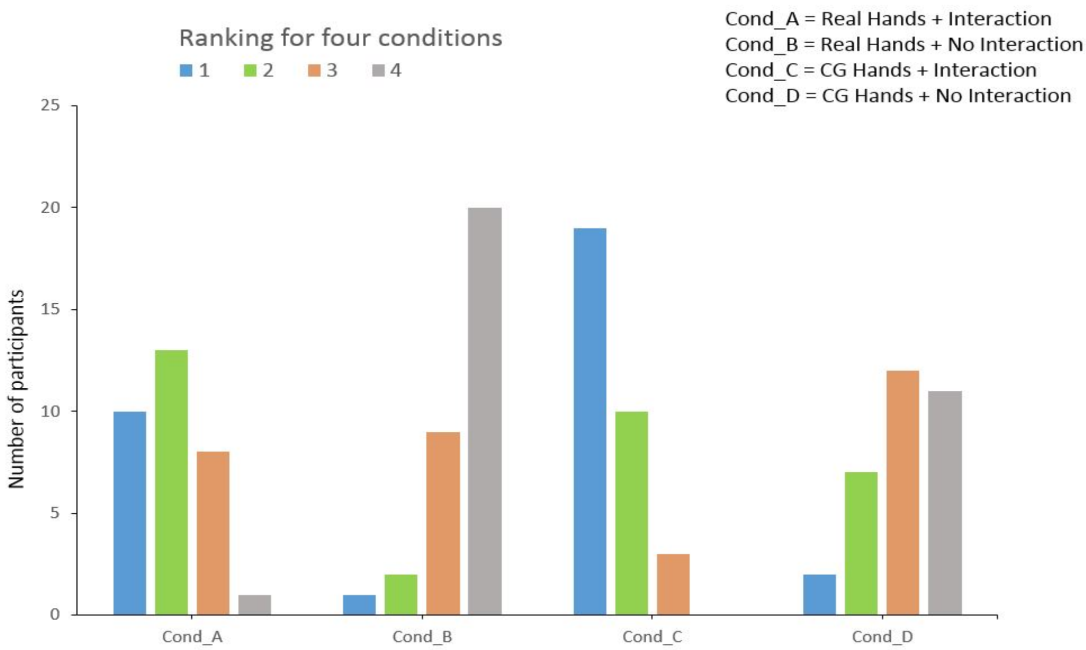


FIGURE 6.9: Conditions ranking bar graph

Ranks

	Mean Rank
Cond_A	2.00
Cond_B	3.50
Cond_C	1.50
Cond_D	3.00

Test Statistics[a]

N	32
Chi-Square	48.000
df	3
Asymp. Sig.	.000

a. Friedman Test

Figure 6.10: Friedman test result for ranking data

Test Statistics[a]

	Cond_B - Cond_A	Cond_C - Cond_A	Cond_D - Cond_A	Cond_C - Cond_B	Cond_D - Cond_B
Z	-4.580[b]	-2.183[c]	-2.868[b]	-4.614[c]	-1.999[c]
Asymp. Sig. (2-tailed)	.000	.029	.004	.000	.046

Test Statistics[a]

	Cond_D - Cond_C
Z	-4.580[b]
Asymp. Sig. (2-tailed)	.000

a. Wilcoxon Signed Ranks Test

b. Based on negative ranks.

c. Based on positive ranks.

Figure 6.11: Wilcoxon signed rank test, post hoc analysis

Participants were also asked whether they would like to see representation of their hands in a 360 VR movie and their response on Likert-scale is in Fig. 6.12.

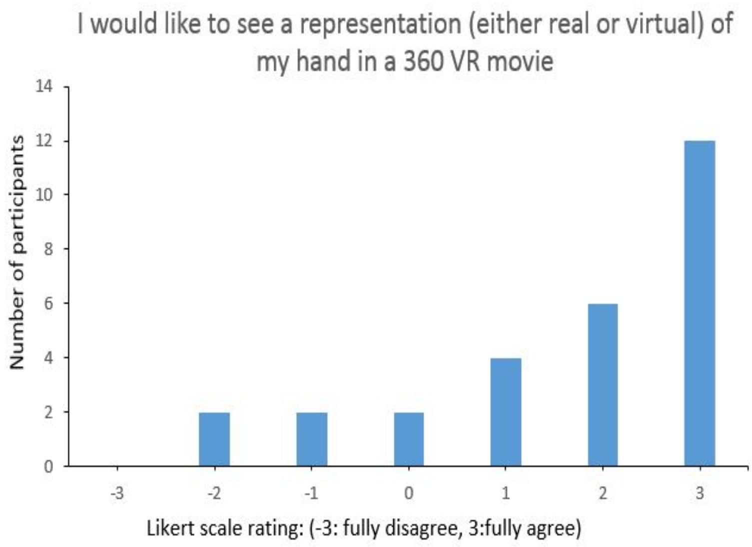


Figure 6.12: Participants response to seeing hands in a 360 VR movie

Using a seven point Likert-scale, we also asked about their inclination for the interaction system ("I would like to use my hands for interaction in a 360 VR movie."), and the response for the question is illustrated in Fig. 6.13.

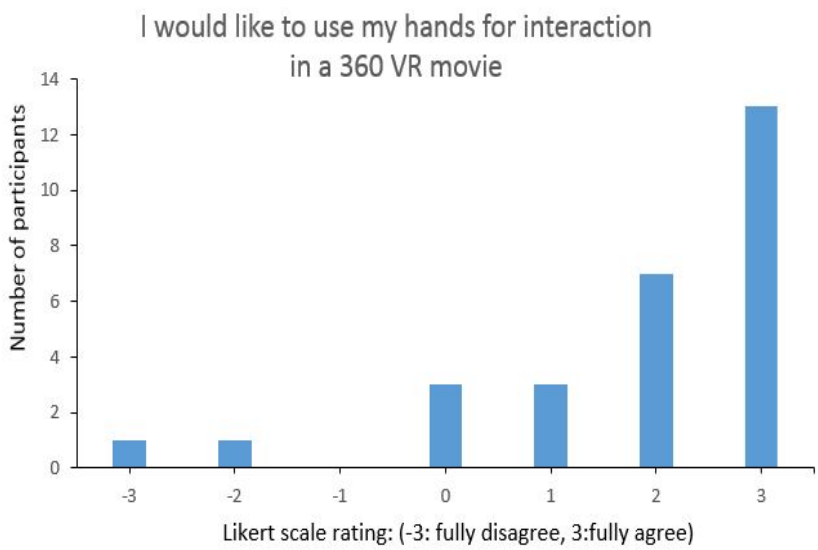


Figure 6.13: Participants response to using hands for interaction in a 360 VR movie

6.2.4 Quantitative Measures Summary

From the quantitative data, having interaction in a 360 movie had a significant effect on two presence components, general presence and involvement. Interaction comparing to no interaction led to higher general presence and involvement score, as shown by box-plots in Fig. 6.2 and Fig. 6.4. It was also found that interaction has a significant effect on sense of agency and ownership.

For type of hands (real hands or CG hands) the data showed significant difference for two presence components, realism and involvement. Real hands had a higher mean score than CG hands, as shown in Fig. 6.6 and Fig. 6.4. For the sense of ownership measured by a questionnaire, type of hands also had a significant effect. Involvement, one of the four components of presence, also showed an interaction effect between the two IVs. A line graph is drawn in Fig. 6.5 to further understand the interaction effect. The line graph shows that with interaction in a 360 movie, real hands have higher mean score than CG hands when there was interaction was allowed, and almost the same scores when no interaction was allowed. Post-experiment questionnaire analysis also showed significant results between the four conditions for ranking data, and that the participants would like to see their hands in a 360 VR movie and be able to interact with the surrounding.

6.3 Qualitative Measures

To further understand the rationale for the choices participants made during the experiment, qualitative measure were also recorded in the post-experiment questionnaire and a short interview session after the experiment.

6.3.1 Post-experiment Questionnaire

Participants were asked to give their reason for the ranking of the four conditions. Here are some of the responses received from the participants.

1. “Because the animation was clean for virtual hands, and it is the condition where it feels ’real’ compared to others.”
2. “Having my own hands made it feel more real. They would move the same way they were moving in the real world.”
3. “The interaction made better sense of being inside the virtual world and visual quality of virtual hand made it better.”

4. "The virtual hands seems to provide a more realistic view compared to the real hands."

5. "I prefer real hands as I can see my own hands, control them and see them as I move them in real world. I am also able to throw seeds in the VR."

6. "The virtual hands appeared clearly and being able to interact made the experience more immersive."

7. "Real hands because I can intertwine my hands, wave my hands, cross my hands, and it is an accurate depiction. All others had same form of difficulty in responding to directions."

8. "The movement of real hands match better, with interaction it feels more interesting"

9. "Had more control than virtual hands. Interaction makes it more realistic."

10. "Virtual hands are more clear, and I think with interaction it is more interesting than just looking around. The real hands were noisy comparing to the virtual hands so I prefer the virtual hands."

11. "Although slightly glitchy, the virtual hands couldn't pick-up every detail of my movements, they fitted better in the virtual world. The shadows on my real hands did not match the virtual environment, and the definition was not very good."

Participants were also asked about any issues experienced during the experiment. Here are some of the comments.

1. "The virtual hand was not aligning properly with my hand. The real hand was low resolution and low quality."

2. "The resolution of real hands is poor. User's position in the movie is a little bit confused like flying in the sky rather than in the hot air balloon."

3. "Besides having to learn that the real hands would disappear from view beyond a certain point."

4. "I would just like to have more interactions. If I would be able to wave my hand and get a response from the other balloon and throw the food to the birds I would get a better sensation of being there."

5. "The action of virtual hands does not react as fast as my real hands. The colour is a bit solid for virtual which makes me feel a bit unreal."

6. "I felt as if the interaction was a bit limited."

Participants were also asked if they have any further suggestions on the experiment. Here are some of the responses.

1. "I wish I was able to interact more with the environment."

2. "The virtual hand representation was fantastic and for the most part reacted the way I expected it to. Representation of the real hands appears to be limited by camera FOV and resolution."

3. "More interactions would be interesting."

4. "I think if virtual hands could have more accurate fingers they would be best."

5. "I suggest that you should use a better camera for the real hand case."

6. "Sometimes, the virtual hand move in really crazy ways, which is quite funny"

7. "The realism was decreased when the birds were flying backwards"

8. "It would be better if we can see our body and feed when we look down"

9. "For virtual hands when fingers touch each other, it feels strange. Crossover, for example, thumb touching the forefinger are not properly represented in VR. The touch feeling makes me prefer noise over fidelity because when you touch your other finger and virtual hand doesn't show that it feels strange."

10. "It's good to have interaction with the movie. I would like more different interactions."

6.3.2 Interview

After collecting the post-experiment questionnaire, the participants who preferred the CG or virtual hands in the questionnaire were asked what would they prefer if the real hands are rendered at the same quality as virtual hands. Five of the participants responded with the real hands and the reason they gave are as follows.

1. "The real hands can follow the gesture when you clap or close two fingers or other similar action, it's always there."
2. "Real hands are good for haptics when your one hand touches another and you can still see the visual output that is cool. But technical limitations make me choose virtual over real hands."
3. "I will prefer real hands if they have better visual quality and don't disappear when you extend them."

There were still three participants who preferred using CG or virtual hands because of the reason as follows.

1. "The virtual hand felt like it matched the environment and belonged there."
2. "The fact that the virtual hands were animated helped with the immersion"

6.3.3 Qualitative Measures Summary

From the qualitative data analysis, it was found that most of the participants preferred interaction over no interaction as they felt more immersed with interactions. For type of hands, more participants preferred CG/virtual hands over real hands because of the visual quality of the CG hands. While few participants preferred real hands over CG/virtual hands, as they felt that the real hands were following the hand movement better than the virtual hands specially for crisscrossed or intertwined hands. Intertwined hands also have a haptics element which was highlighted by one of the participants who chose real hands.

The participants also gave their feedback on the problems they experienced during the experiment. The problems highlighted were mainly, visual quality of the real hands, the loss of tracking for virtual hands, the limited field of view for real hands and the limited amount of interaction available in the virtual space. The participants also suggested improvements for the experiment which were similar to the highlighted problems. Some of the suggested improvements were:

- Using higher resolution camera with larger field of view for the real hands.

- Virtual hands with improved tracking so that the hands does not disappear when the hands in real world overlap.
- Include more types of interactions in virtual space as currently, there is only feeding interaction.
- Use the whole avatar or real body in virtual space so when user looks down s/he can see his whole body instead of just hands.

Chapter 7

Discussion

This chapter discusses the results found in the user study. It also explores possible explanations of the results and their relationship with previous research work. It also explains the limitations of the study.

7.1 Study Results

7.1.1 Sense of Presence

The results of the IPQ questionnaire are divided into four components (general presence, spatial presence, involvement and experienced realism) as explained in Chapter 5. For the hand gesture-based interaction, there were significant effects on the general presence and involvement. The descriptive statistics of the general presence and involvement revealed that interaction had a positive effect on both of the measures. The hand appearance had significant effects on the experienced realism and involvement. The descriptive statistics showed that the real hands had higher realism and involvement than the virtual hands. Both of these results support our hypotheses, H1 (the interaction increases the sense of presence) and H2 (the real hands increases the sense of presence). There was no significant effect found for the spatial presence. This indicates that the hand appearance or the hand gesture-based interaction does not affect the spatial presence.

The involvement results showed an interaction effect between the two IVs which is shown in Fig. 6.5. The interaction effect indicates that the user felt more involvement with the real hands when there was interactivity with the virtual environment, but the involvement was same for the real hands and the CG hands when there was no interactivity.

The overall IPQ questionnaire results showed no significant effect. This could be because of the spatial presence component which also did not have any significant effect. The spatial presence is the sense of being physically there. The spatial presence might be breaking because of the stitching in the 360 video. The stitched part of the video is right below the user. When the user feeds the bird, they come to collect the food. The user looks down to see the bird; he also sees the stitched part which might be causing the break in spatial presence. The break in presence might also be because of not showing the full body of the user. So when the user looks down and does not see the body, it causes a break in the spatial presence. This is also indicated in the qualitative feedback by the participant UE_22 who stated, “It would be better if we can see our body and feed when we look down”. Slater [54] also points to this: ”In physical reality and first-order virtual reality, there is something very simple that you can do to physically establish your presence. Look down, and you will see your body, or see parts of it”. However, a future study should investigate the causes of the break in spatial presence and whether that would result in the overall increase in presence. The overall results indicate no significant effect, so the hypotheses H1 and H2 are not supported.

7.1.2 Sense of Embodiment

The results of the embodiment questionnaire are divided into two parts: agency and ownership. For the hand gesture-based interaction, there were significant effects on the agency and ownership. The descriptive statistics of the agency and ownership indicate that having interaction results in higher agency and ownership than without interaction. This confirms the hypotheses, H3 (the interaction increases the sense of agency) and H5 (the interaction increases the sense of ownership). For the hand appearance, there was a significant effect on the ownership. The descriptive statistics shows that the real hands had higher ownership than the virtual hands; hence, the hypothesis, H6 (the real hands increases the sense of ownership) is confirmed. The hand appearance did not affect the sense of agency; therefore, the hypothesis, H4 (the real hands increases the sense of agency) is not supported.

The results showed for the sense of embodiment are in line with Lin and Jörg [14] results. They also found that the sense of agency is not affected by the hand appearance, and the sense of ownership increases with the hand appearance. The virtual hand models used by Lin and Jörg is shown in Fig. 7.1.

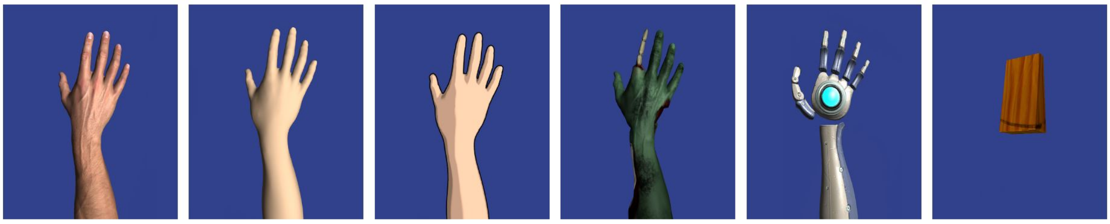


FIGURE 7.1: Hand models used by Lin and Jörg, from left to right: realistic hand, toony hand, very toony hand, zombie hand, robot hand, wooden block.[14]

Another work by Argelaguet et al. [15] obtained the same results for the sense of ownership, but different results for the sense of agency. They found that the ownership increases with the realistic hands. However, they found that realistic hands have a lower sense of agency than non-realistic hands. The virtual hand models used by Argelaguet et al. are shown in Fig. 7.1.

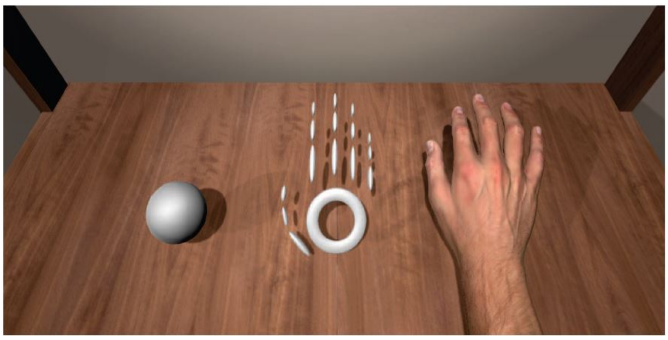


FIGURE 7.2: Hand models used by Argelaguet et al.: Abstract hands(left), iconic hands (center) and realistic virtual hands (right).[15]

7.1.3 Overall Discussion

From the experiment results, we can now answer our research questions. *Does hand gesture-based interaction in a cinematic virtual environment improve a users sense of presence and embodiment?*. The results show that the hand gesture-based interaction improves the sense of embodiment, but has no effect on the sense of presence. So the answer to our first research question is partly yes and partly no. The reason why there was no significant effect on the sense of presence by the hand gesture-based interaction might be the spatial presence as identified in the earlier section 7.1.1. This should be further investigated in the future study.

Does the type of hand appearance (actual hands or CG hands) in a cinematic virtual environment affect the sense of presence and embodiment? The results showed that the hand appearance partly improves the sense of embodiment. It only effects the sense of ownership and has no effect on the sense of agency. For the sense of presence, the hand appearance also has no effect. So the answer to our question is mostly no but partly yes in terms of the sense of ownership. For the sense of presence, no significant found might due to the appearance of the real hands which were not as good as the actual hands. The real hands were rendered using a point cloud. Even though they elicit a stronger sense of ownership but the visual fidelity is not enough for creating the sense of presence. Our hand appearance results are also similar to Lok et al. [8]. They also found that visual fidelity does not increase the sense of presence. However, like our real hand representation their real hand representation is also not as good as the actual hands. Their segmentation still shows the real world surrounding the hands, as can be seen in Fig. 2.9. The effect of hand appearance is probably going to be subtle, but it should be further studied. For the sense of agency, our results corroborate the past research [14] but are also in contrast with the Argelaguet et al. [15]. The hand appearance probably would not effect the sense of agency. So the future work on agency should be focused on the interaction part.

7.2 Limitations

While developing the prototype, we came across some of the limitations in the system which should be considered by the future researcher. There were also some limitations identified by the participants during the study which could be helpful for the future researchers.

- The real hands were visualised using a point cloud which was discontinuous at places. The fidelity was enough for the user to perceive as his own hand but was still not clear enough as high quality image of the real hands. Probably for the future study, render a 3D model of the hand from the captured data. The real-time rendering should not lower the frame rate inside the VR as that will cause motion-sickness. One solution could be writing a unity shader to process the data on GPU.

- The SoftKinetic camera has a smaller field of view which also decreases immersion. In the future when a RGBD camera with better field of view is available. The study should be validated again. The results might be different as the field of view could be a limiting factor. The field of view can be improved with the current SoftKinetic camera by using an external lens; however, it could interfere with the cameras time-of-flight measurement. If it improves the field of view without interfering with the camera's system, it would be a good solution.

- For the virtual hand models when the fingers are all together, they will overlap with each other. This cause break in presence. It is caused due to the limitation of the hand tracking. For the real hands, it was not the case as the RGB image of the hands were shown. So when the fingers are together, they will be shown as together.

- The virtual hand model were of half arm length. When the user fully extended their hand, the end part of the arm can be seen. This could have caused break in presence. For future study, a full arm length hand model should be used.

- With the current leap motion hand tracking the interaction with the small virtual objects is not possible so the user study which involves small virtual objects should consider another tracking solution.

- The 360 video with better stitching should be used. The point of view of the user should also be considered specially for interaction based scenarios. Because if the user is constantly seeing their hand going through objects or people shown in the video that will decrease the sense of presence. Also the user point of view in the 360 video should be chosen as such that the interaction is plausible with the video.

Chapter 8

Conclusion and Future Work

This chapter briefly summarises the thesis and summarises the conclusions drawn from the results. It also identifies possible future areas of research.

8.1 Conclusion

In this thesis, we studied the effect of hand gesture-based interaction on the sense of presence and embodiment. We also investigated the effect of hand appearance on the sense of presence and embodiment. For hand appearance, we had two types of hand, real hand and virtual hand. For the virtual hand, it is native to the virtual world so we have to drive it based on the real hand information which is obtained through hand tracking. But for the real hand, it needs to be include as well as tracked. To include the real hand, it has to be captured by a camera, then segmented out of the image, and finally rendered inside the virtual world. The prototype developed in the project captured the real hand and allowed users to interact with the virtual world. The prototype has the Leap Motion as a hand tracker and the SoftKinetic camera for hand visualisation.

The prototype was used for a user experiment in which we studied the hand-based interaction in a 360 VR movie. We investigated the effect of interaction and hand appearance on the sense of presence and embodiment. Both interaction and hand appearance had two levels in the study. For interaction, the two levels were: with interaction, and without interaction. For the hand appearance, the two levels were: the real hands, and the virtual hands. There were 32 participants in the user study. The study results showed that the user had higher sense of embodiment with interaction and using the real hands. The result also indicated that the interaction and hand appearance has no effect on the sense of presence. The sense of presence was measured using the IPQ questionnaire. The IPQ components were also analyses which indicated that both the interaction and hand appearance have no effect on the spatial presence. Based on that we suggested that the VR components which creates the sense of spatial presence should be further studied and analysed. There are also limitations to the study which are provided in the previous chapter.

8.2 Future Work

For future work there are several possibilities. Some of them are identified in the limitation section of the Chapter 7. Other possibilities are as follows.

- In section 3.2, we identified four quadrants (observant active, observant passive, participant active, and participant active) of interaction in a VR video. From the four quadrants, we experimented on the participant passive quadrant and found significant effects on the sense of presence. In the future study, the participant active quadrant can be experimented. The open source Henry's animation can be used for the experiment.

- A user study comparing offline rendered real hand texture with the real-time rendered hand textures can also be conducted. The study should measure the effect on sense of ownership. The study will reveal the subtle difference between the two. We hypothesise that the real-time rendered hand textures will have the higher sense of ownership.

- A user study can be conducted to identify different types of virtual hand models for a particular cinematic environments. One hand model should be rendered with a costume or some resemblance to the cinematic environment and can be compared with other hand models which are not specific to that cinematic experience. The sense of embodiment should be used as measures.

- The user study in this thesis limited interaction to the manipulation. Other interaction modes (such as navigation) and their effect on the sense of presence and embodiment can also be studied.

- On the technical aspect, the hand tracker can be combined with the full body tracking. Study the effect of a fully tracked avatar on the sense of presence and embodiment. Another experiment could be rendering the full user body that is tracked by the full body tracker and compare it with an avatar body in the cinematic VR environment.

Bibliography

[1] M. Carrozzino, C. Lorenzini, M. Duguleana, C. Evangelista, R. Brondi, F. Tecchia, and M. Bergamasco, "An immersive vr experience to learn the craft of printmaking," in *International Conference on Augmented Reality, Virtual Reality and Computer Graphics.* Springer, 2016, pp. 378–389.

[2] A. D. Cheok, W. Weihua, X. Yang, S. Prince, F. S. Wan, M. Billinghurst, and H. Kato, "Interactive theatre experience in embodied+ wearable mixed reality space," in *Proceedings of the 1st International Symposium on Mixed and Augmented Reality.* IEEE Computer Society, 2002, p. 59.

[3] F. V. Studio, *Interstellar Virtual Reality Experience*, (Accessed: March 14, 2017). [Online]. Available: http://framestorevr.com/interstellar-experience/

[4] P. Milgram and F. Kishino, "A taxonomy of mixed reality visual displays," *IEICE TRANSACTIONS on Information and Systems*, vol. 77, no. 12, pp. 1321–1329, 1994.

[5] M. McGill, D. Boland, R. Murray-Smith, and S. Brewster, "A dose of reality: overcoming usability challenges in vr head-mounted displays," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems.* ACM, 2015, pp. 2143–2152.

[6] P. Budhiraja, R. Sodhi, B. Jones, K. Karsch, B. Bailey, and D. Forsyth, "Where's my drink? enabling peripheral real world interactions while using hmds," *arXiv preprint arXiv:1502.04744*, 2015.

[7] F. Tecchia, G. Avveduto, R. Brondi, M. Carrozzino, M. Bergamasco, and L. Alem, "I'm in vr!: using your own hands in a fully immersive mr system," in *Proceedings*

of the 20th ACM Symposium on Virtual Reality Software and Technology. ACM, 2014, pp. 73–76.

[8] B. Lok, S. Naik, M. Whitton, and F. Brooks, "Effects of handling real objects and avatar fidelity on cognitive task performance in virtual environments," in *Virtual Reality, 2003. Proceedings. IEEE.* IEEE, 2003, pp. 125–132.

[9] J. Y. Chen, "Augmenting immersive cinematic experience/scene with user body visualization," Thesis, 2016.

[10] TechCrunch, *Redefining The Axiom Of Story: The VR And 360 Video Complex*, January 14, 2016. [Online]. Available: https://techcrunch.com/2016/01/14/redefining-the-axiom-of-story-the-vr-and-360-video-complex/

[11] A. Tagliasacchi, M. Schröder, A. Tkach, S. Bouaziz, M. Botsch, and M. Pauly, "Robust articulated-icp for real-time hand tracking," in *Proceedings of the Eurographics Symposium on Geometry Processing*, ser. SGP '15. Aire-la-Ville, Switzerland, Switzerland: Eurographics Association, 2015, pp. 101–114.

[12] H. Bai, L. Gao, J. El-Sana, and M. Billinghurst, "Free-hand interaction for handheld augmented reality using an rgb-depth camera," in *SIGGRAPH Asia 2013 Symposium on Mobile Graphics and Interactive Applications*, ser. SA '13. New York, NY, USA: ACM, 2013, pp. 22:1–22:4.

[13] L. Motion, *Hand Tracking in Virtual Reality*, (Accessed: February 9, 2017). [Online]. Available: https://developer.leapmotion.com/documentation/unity/unity/Unity_VR.html

[14] L. Lin and S. Jörg, "Need a hand?: How appearance affects the virtual hand illusion," in *Proceedings of the ACM Symposium on Applied Perception*, ser. SAP '16. New York, NY, USA: ACM, 2016, pp. 69–76.

[15] F. Argelaguet, L. Hoyet, M. Trico, and A. Lecuyer, "The role of interaction in virtual embodiment: Effects of the virtual hand representation," in *2016 IEEE Virtual Reality (VR)*, March 2016, pp. 3–10.

[16] S. S. Rautaray and A. Agrawal, "Vision based hand gesture recognition for human computer interaction: a survey," *Artificial Intelligence Review*, vol. 43, no. 1, pp. 1–54, 2015.

[17] M. Mine *et al.*, “Virtual environment interaction techniques,” *UNC Chapel Hill computer science technical report TR95-018*, pp. 507 248–2, 1995.

[18] O. Hilliges, D. Kim, S. Izadi, M. Weiss, and A. Wilson, “Holodesk: direct 3d interactions with a situated see-through display,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, 2012, pp. 2421–2430.

[19] J. Wang and R. W. Lindeman, “Comparing isometric and elastic surfboard interfaces for leaning-based travel in 3d virtual environments,” in *3D User Interfaces (3DUI), 2012 IEEE Symposium on.* IEEE, 2012, pp. 31–38.

[20] C. A. Zanbaka, B. C. Lok, S. V. Babu, A. C. Ulinski, and L. F. Hodges, “Comparison of path visualizations and cognitive measures relative to travel technique in a virtual environment,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 11, no. 6, pp. 694–705, 2005.

[21] R. Tenmoku, R. Ichikari, F. Shibata, A. Kimura, and H. Tamura, “Design and prototype implementation of mr pre-visualization workflow,” in *DVD-ROM Proc. Int. Workshop on Mixed Reality Technology for Filmmaking*, 2006, pp. 1–7.

[22] R. Ichikari, K. Kikuchi, W. Toishita, R. Tenmoku, F. Shibata, and H. Tamura, “On-site real-time 3d match move for mr-based previsualization with relighting,” in *ACM SIGGRAPH 2010 Talks*, ser. SIGGRAPH ’10. New York, NY, USA: ACM, 2010, pp. 36:1–36:1.

[23] C. Nolan, *Interstellar*, 2014.

[24] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino, “Augmented reality: a class of displays on the reality-virtuality continuum,” pp. 282–292, 1995.

[25] H. Regenbrecht, M. Haller, J. Hauber, and M. Billinghurst, “Carpeno: interfacing remote collaborative virtual environments with table-top interaction,” *Virtual Reality*, vol. 10, no. 2, pp. 95–107, 2006.

[26] H. Maeda, T. Tanikawa, J. Yamashita, K. Hirota, and M. Hirose, “Real world video avatar: transmission and presentation of human figure,” pp. 237–238, March 2004.

[27] P. J. Metzger, “Adding reality to the virtual,” pp. 7–13, Sep 1993.

[28] T. Günther, I. S. Franke, and R. Groh, "Aughanded virtuality-the hands in the virtual environment," in *3D User Interfaces (3DUI), 2015 IEEE Symposium on.* IEEE, 2015, pp. 157–158.

[29] G. Bruder, F. Steinicke, K. Rothaus, and K. Hinrichs, "Enhancing presence in head-mounted display environments by visual body feedback using head-mounted cameras," in *CyberWorlds, 2009. CW'09. International Conference on.* IEEE, 2009, pp. 43–50.

[30] T. Ha, S. Feiner, and W. Woo, "Wearhand: Head-worn, rgb-d camera-based, bare-hand user interface with visually enhanced depth perception," in *Mixed and Augmented Reality (ISMAR), 2014 IEEE International Symposium on.* IEEE, 2014, pp. 219–228.

[31] M. Usoh, E. Catena, S. Arman, and M. Slater, "Using presence questionnaires in reality," *Presence: Teleoperators and Virtual Environments*, vol. 9, no. 5, pp. 497–503, 2000.

[32] T. Schubert, F. Friedmann, and H. Regenbrecht, "The experience of presence: Factor analytic insights," *Presence: Teleoperators and virtual environments*, vol. 10, no. 3, pp. 266–281, 2001.

[33] F. Bach, H. Çakmak, and H. Maaß, "Vision-based hand representation and intuitive virtual object manipulation in mixed reality," *Biomedical Engineering/Biomedizinische Technik*, vol. 57, no. SI-1 Track-B, pp. 462–465, 2012.

[34] L. P. Fiore and V. Interrante, "Towards achieving robust video selfavatars under flexible environment conditions," *International Journal of Virtual Reality*, vol. 11, no. 3, pp. 33–41, 2012.

[35] G. Park, T. Ha, and W. Woo, "Hand tracking with a near-range depth camera for virtual object manipulation in an wearable augmented reality," in *Proceedings, Part I, of the 6th International Conference on Virtual, Augmented and Mixed Reality. Designing and Developing Virtual and Augmented Environments - Volume 8525.* New York, NY, USA: Springer-Verlag New York, Inc., 2014, pp. 396–405.

[36] M. Slater and S. Wilbur, "A framework for immersive virtual environments (five): Speculations on the role of presence in virtual environments," *Presence: Teleoperators and virtual environments*, vol. 6, no. 6, pp. 603–616, 1997.

[37] M. Usoh, K. Arthur, M. C. Whitton, R. Bastos, A. Steed, M. Slater, and F. P. Brooks Jr, "Walking¿ walking-in-place¿ flying, in virtual environments," in *Proceedings of the 26th annual conference on Computer graphics and interactive techniques.* ACM Press/Addison-Wesley Publishing Co., 1999, pp. 359–364.

[38] K. Kilteni, R. Groten, and M. Slater, "The sense of embodiment in virtual reality," *Presence: Teleoperators and Virtual Environments*, vol. 21, no. 4, pp. 373–387, 2012.

[39] T. Bayne and E. Pacherie, "Narrators and comparators: the architecture of agentive self-awareness," *Synthese*, vol. 159, no. 3, pp. 475–491, 2007.

[40] M. Tsakiris, G. Prabhu, and P. Haggard, "Having a body versus moving your body: How agency structures body-ownership," *Consciousness and cognition*, vol. 15, no. 2, pp. 423–432, 2006.

[41] M. V. Sanchez-Vives, B. Spanlang, A. Frisoli, M. Bergamasco, and M. Slater, "Virtual hand illusion induced by visuomotor correlations," *PloS one*, vol. 5, no. 4, p. e10381, 2010.

[42] T. C. Peck, S. Seinfeld, S. M. Aglioti, and M. Slater, "Putting yourself in the skin of a black avatar reduces implicit racial bias," *Consciousness and cognition*, vol. 22, no. 3, pp. 779–787, 2013.

[43] N. Y. Times, "Interactive filmmakers hope to make a comeback," August 17, 1998 (Accessed: March 8, 2017). [Online]. Available: http://partners.nytimes.com/library/tech/98/08/cyber/articles/17dvd.html

[44] B. Bejan, *I'm Your Man*, 1992.

[45] L. Motion, *Leap Motion for Mac and PC*, (Accessed: February 10, 2017). [Online]. Available: https://www.leapmotion.com/product/desktop#107

[46] JaeYunLee, *Easy Movie Texture*, (Accessed: February 15, 2017). [Online]. Available: https://www.assetstore.unity3d.com/en/#!/content/10032

[47] SoftKinetic, *DS325 DataSheet*, (Accessed: February 8, 2017). [Online]. Available: http://www.softkinetic.com/Portals/0/Download/WEB_20120907_SK_DS325_Datasheet_V2.1.pdf

[48] P. J. Besl and N. D. McKay, "A method for registration of 3-d shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 239–256, Feb. 1992.

[49] J. van Baren and W. IJsselsteijn, *Measuring Presence: A Guide to Current Measurement Approaches*, (Accessed: February 12, 2017). [Online]. Available: http://www8.informatik.umu.se/~jwworth/PresenceMeasurement.pdf

[50] H. Regenbrecht and T. Schubert, "Real and illusory interactions enhance presence in virtual environments," *Presence: Teleoper. Virtual Environ.*, vol. 11, no. 4, pp. 425–434, Aug. 2002.

[51] T. W. Schubert, "The sense of presence in virtual environments:," *Zeitschrift fr Medienpsychologie*, vol. 15, no. 2, pp. 69–71, 2003.

[52] T. Schubert, F. Friedmann, and H. Regenbrecht, *Embodied Presence in Virtual Environments*. London: Springer London, 1999, pp. 269–278.

[53] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins, "The aligned rank transform for nonparametric factorial analyses using only anova procedures," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '11. New York, NY, USA: ACM, 2011, pp. 143–146.

[54] M. Slater, "Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 364, no. 1535, pp. 3549–3557, 2009.

Appendix A

Appendix A: Information Sheet and Consent Form

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

HUMAN ETHICS COMMITTEE

Secretary, Rebecca Robinson
Telephone: +64 03 369 4588, Extn 94588
Email: human-ethics@canterbury.ac.nz

Ref: HEC 2016/73/LR

21 December 2016

Humayun Khan
HIT Lab NZ
UNIVERSITY OF CANTERBURY

Dear Humayun

Thank you for submitting your low risk application to the Human Ethics Committee for the research proposal titled "Hand Gestures Based Interaction in an Immersive Cinematic Environment".

I am pleased to advise that the application has been reviewed and approved.

Please note that this approval is subject to the incorporation of the amendments you have provided in your email of 20th December 2016.

With best wishes for your project.

Yours sincerely

pp. R. Robinson

Associate Professor Jane Maidment
Chair, Human Ethics Committee

HIT Lab NZ
Telephone: +64 364 2349
Email: humayun.khan@pg.canterbury.ac.nz
12.12.2016

HITLabNZ
Human Interface Technology Lab New Zealand
Hangarau Tangata, Tangata Hangarau

PARTICIPANT INFORMATION SHEET

RESEARCH STUDY: **Hand gestures based interaction in an Immersive Cinematic Environment**

RESEARCHERS: **Humayun Khan, Prof. Rob Lindeman, Dr. Gun Lee, Prof. Mark Billinghurst**

INTRODUCTION

You are invited to take part in a cinematic experience research study. Before you decide to be part of this study, you need to understand the risks and benefits. This information sheet provides information about the research study. The researcher will be available to answer your questions and provide further explanations. If you agree to take part in the research study, you will be asked to sign the consent form.

PURPOSE

The purpose of this study is to study the effect of hand gestures based interaction on user experience in a cinematic environment.

PROCEDURE

The study will follow the procedure outlined as below:

1. The participant reads information sheet and signs the consent form.
2. The participant answers to a pre-experiment questionnaire about demographic information and their previous experience with using HMD.
3. The researcher explains the study setup and experimental tasks for the participant to perform during the study.
4. The participant performs the experimental tasks including:
   - Wearing the Oculus Rift(HMD) and pilot testing with mounted cameras
   - Perform the given task, which may include watching a short movie, viewing one's own hands inside the movie, and interacting with the objects spheres and cubes in the movie.
   - Rate personal thoughts and feelings based on each task by answering a questionnaire.
5. The participant answers a post-experiment questionnaire asking for feedback on the overall study.

The whole procedure will take approximately 40 minutes.

RISKS/DISCOMFORTS

Risks are dependent on individuals in this study. As you will be wearing a head-mounted display (HMD) a number of times, you might feel some form of nausea or giddiness. Please let the researcher know if you feel nauseous or any discomfort. Being still and with eyes closed could help prevent causing the nausea.

CONFIDENTIALITY

All data obtained from participants will be kept confidential. In publications (e.g. Thesis, a public document which will be available through the UC Library), we will mainly report the results in an aggregate format: individual scores will be combined into a cumulative score. In case of reporting the participant's quotes from the questionnaires, we will keep the source anonymous. Video of the experiment will be recorded for analysis purposes. The recorded video will be only of what participants can see through the Oculus Rift. There will be no recordings of the participant's face, maintaining the anonymity of the participants. All recordings will be securely stored, and no one other than the researchers will have access to them. The data will be kept securely for a minimum period of 5 years and will be destroyed upon completion of the research project.

PARTICIPATION

Participation in this research study is completely voluntary. You have the right to withdraw at any time or refuse to participate entirely.

COMPENSATION

Upon completion, the participant will receive a $10 gift voucher.

RESULTS

Please indicate to the researcher on the consent form if you would like to receive a copy of the summary of results of the project.

APPROVAL OF THIS STUDY

This project has been reviewed and approved by the University of Canterbury Human Ethics Committee, and participants should address any complaints to The Chair, Human Ethics Committee, University of Canterbury, Private Bag 4800, Christchurch (human-ethics@canterbury.ac.nz).

QUESTIONS

If you have any questions regarding this study, please contact:
Humayun Khan (humayun.khan@pg.canterbury.ac.nz)

Please take this information sheet with you when you leave.

PARTICIPANT CONSENT FORM

RESEARCH STUDY: **Augmenting Immersive Cinematic Experience/Scene with User Body Visualisation**

RESEARCHER: **Humayun Khan (humayun.khan@pg.canterbury.ac.nz)**

SUPERVISORS: **Prof. Rob Lindeman (gogo@hitlabnz.org), Dr. Gun Lee (gun.lee@canterbury.ac.nz), Prof. Mark Billinghurst (mark.billinghurst@canterbury.ac.nz)**

I have been given a full explanation of this project and have had the opportunity to ask questions.
I understand what is required of me if I agree to take part in the research.

I understand that participation is voluntary and I may withdraw at any time without penalty. Withdrawal of participation will also include the withdrawal of any information I have provided should this remain practically achievable.

I understand that any information or opinions I provide will be kept confidential to the researcher and the administrators of the research project and that any published or reported results will not identify the participants. I understand that a thesis is a public document and will be available through the UC Library.

I understand that whatever I can see through the head-mounted display will be recorded in video form.

I understand that all data collected for the study will be kept in locked and secure facilities and/or in password protected electronic form and will be destroyed after five years.

I understand the risks associated with taking part and how they will be managed.

I understand that I am able to receive a report on the findings of the study by contacting the researcher at the conclusion of the project.

I understand that I can contact the researchers or supervisors listed above for further information. If I have any complaints, I can contact the Chair of the University of Canterbury Human Ethics Committee, Private Bag 4800, Christchurch (humanethics@canterbury.ac.nz)

I would like a summary of the results of the project

By signing below, I agree to participate in this research project, and I authorize recordings or other materials taken from this study used for scientific purposes, and I consent to publication of the results of the study.

____________________________ __________________ ________

Participant (Print name) Signature Date

Email *(for report of findings, if applicable)*: ____________________________

Appendix B

Appendix B: Questionnaires

Pre-experiment Questionnaire

To be filled by researcher:

Participant No.	Hand Model No.

Q1 How old are you?

______ years old.

Q2 What is your gender?

❍ Male
❍ Female
❍ Other

Q3 Have you used Head-Mounted Displays before? (For example, Oculus Rift, HTC Vive or Samsung Gear VR)

❍ Never
❍ Few times a year
❍ Few times a month
❍ Few times a week
❍ Everyday

Q4 Which Head-Mounted Displays have you used before? Tick all that apply

☐ Oculus Rift
☐ HTC Vive
☐ Sony PSVR
☐ Samsung Gear VR
☐ Microsoft HoloLens
☐ Google Cardboard
☐ Others (______________________________)

Q5 What do you prefer for entertainment? Tick all that apply

☐ Watching movies at home
☐ Watching movies in cinema
☐ Computer gaming
☐ Console gaming (e.g. Play Station, Xbox, Nintendo etc)
☐ Mobile gaming
☐ Others (______________________________)

Q6 Are you afraid of heights?

❍ Yes
❍ No
❍ Not sure

Per-condition Questionnaire

To be filled by researcher:

Participant No.	Condition No.

Q1 How aware were you of the real world surrounding while navigating in the virtual world? (i.e. sounds, room temperature, other people, etc.)?

extremely aware — moderately aware — not aware at all

O	O	O	O	O	O	O
-3	-2	-1	0	+1	+2	+3

Q2 How real did the virtual world seem to you?

completely real — not real at all

O	O	O	O	O	O	O
-3	-2	-1	0	+1	+2	+3

Q3 I had a sense of acting in the virtual space, rather than operating something from outside.

fully disagree — fully agree

O	O	O	O	O	O	O
-3	-2	-1	0	+1	+2	+3

Q4 How much did your experience in the virtual environment seem consistent with your real world experience?

not consistent — moderately consistent — very consistent

O	O	O	O	O	O	O
-3	-2	-1	0	+1	+2	+3

Q5 How real did the virtual world seem to you?

about as real as an imagined world — indistinguishable from the real world

O	O	O	O	O	O	O
-3	-2	-1	0	+1	+2	+3

Q6 I did not feel present in the virtual space.

did not feel						felt present
O	O	O	O	O	O	O
-3	-2	-1	0	+1	+2	+3

Q7 I was not aware of my real environment.

fully disagree						full agree
O	O	O	O	O	O	O
-3	-2	-1	0	+1	+2	+3

Q8 In the computer generated world I had a sense of "being there".

not at all						very much
O	O	O	O	O	O	O
-3	-2	-1	0	+1	+2	+3

Q9 Somehow I felt that the virtual world surrounded me.

fully disagree						fully agree
O	O	O	O	O	O	O
-3	-2	-1	0	+1	+2	+3

Q10 I felt present in the virtual space.

fully disagree						fully agree
O	O	O	O	O	O	O
-3	-2	-1	0	+1	+2	+3

Q11 I still paid attention to the real environment.

fully disagree						fully agree
O	O	O	O	O	O	O
-3	-2	-1	0	+1	+2	+3

Q12 The virtual world seemed more realistic than the real world.

fully disagree						fully agree
O	O	O	O	O	O	O
-3	-2	-1	0	+1	+2	+3

Q13 I felt like I was just perceiving pictures.

fully disagree						fully agree
O	O	O	O	O	O	O
-3	-2	-1	0	+1	+2	+3

Q14 I was completely captivated by the virtual world.

fully disagree						fully agree
O	O	O	O	O	O	O
-3	-2	-1	0	+1	+2	+3

Q15 I felt as if the virtual representation of the hand moved just like I wanted it to, as if it was obeying my will.

fully disagree						fully agree
O	O	O	O	O	O	O
-3	-2	-1	0	+1	+2	+3

Q16 I expected the virtual representation of the hand to react in the same way as my own hand.

fully disagree						fully agree
O	O	O	O	O	O	O
-3	-2	-1	0	+1	+2	+3

Q17 The interaction was (-3 difficult, 3 easy) to perform.

difficult						easy
O	O	O	O	O	O	O
-3	-2	-1	0	+1	+2	+3

Q18 I felt like I was able to interact with the environment the way I wanted to.

fully disagree						fully agree
O	O	O	O	O	O	O
-3	-2	-1	0	+1	+2	+3

Q19 I felt that the interaction with the environment was realistic.

fully disagree						fully agree
O	O	O	O	O	O	O
-3	-2	-1	0	+1	+2	+3

Q20 I felt as if the virtual representation of the hand was part of my body.

fully disagree						fully agree
O	O	O	O	O	O	O
-3	-2	-1	0	+1	+2	+3

Q21 I felt as if the virtual representation of the hand was someone else's.

fully disagree						fully agree
O	O	O	O	O	O	O
-3	-2	-1	0	+1	+2	+3

Q22 I felt that I was losing the control of my hand when the virtual hand was not responding properly.

fully disagree						fully agree
O	O	O	O	O	O	O
-3	-2	-1	0	+1	+2	+3

Post-experiment Questionnaire

To be filled by researcher:

Participant No.	Hand Model No.

Q1 Which condition do you prefer for watching a 360 movie? Please rank the conditions by writing a number (1:best to 4:worst) inside the boxes beside the options below

☐ Real hands with Interaction
☐ Virtual hands with Interaction
☐ Real hands without Interaction
☐ Virtual hands without Interaction

Q2 Why do you think that the condition you ranked as #1 is the best interaction method for you?

____
____
____
____

Q3 I would like to see a representation (either real or virtual) of my hand in a 360 VR movie.

fully disagree						fully agree
O	O	O	O	O	O	O
-3	-2	-1	0	+1	+2	+3

Q4. Which type of representation of your hand do you prefer?

☐ Real hands
☐ Virtual hands
☐ No preference

Q5 I would like to use my hands for interaction in a 360 VR movie.

fully disagree						fully agree
O	O	O	O	O	O	O
-3	-2	-1	0	+1	+2	+3

Q6 Did you have any issues during the experiment?

Q7 Do you have any other comments on the experiment?